

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 22:34:29 : Search time 32.85 Seconds

(Without alignments) 141.819 Million cell updates/sec

Title: US-09-471-255-2_COPY_15_150

Perfect score: 711
Sequence: 1 IVSLSLCAVFLNQHNSQENK.....DGKYYVYLKDAHADNVRK 136

oring table: BLOSOM62
Gapop 10.0, Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match Length | ID | Description |
|------------|-------|--------------------|--------|-------------|
| 1 | 79 | 11.1 | 315 1 | CALU_MOUSE |
| 2 | 76.5 | 10.8 | 622 1 | PPID_HAEN |
| 3 | 74.5 | 10.5 | 743 1 | BGAL_THERT |
| 4 | 74.5 | 10.5 | 836 1 | UMES_YEAST |
| 5 | 74 | 10.4 | 710 1 | PPCE_HUMAN |
| 6 | 74 | 10.4 | 710 1 | PPCE_PIG |
| 7 | 73.5 | 10.3 | 113 1 | PRML_BRARE |
| 8 | 73 | 10.3 | 417 1 | SENA_APLCA |
| 9 | 73 | 10.3 | 417 1 | Y943_MERJA |
| 10 | 72 | 10.1 | 238 1 | E6_GOSHI |
| 11 | 72 | 10.1 | 517 1 | EAS_DROME |
| 12 | 72 | 10.1 | 1116 1 | SLPH_BACBR |
| 13 | 72 | 10.1 | 1592 1 | GF72_STRDO |
| 14 | 71.5 | 10.1 | 753 1 | NECI_MOUSE |
| 15 | 71.5 | 10.1 | 752 1 | SLAP_ACEKI |
| 16 | 71.5 | 10.1 | 903 1 | ECIC_BOVIN |
| 17 | 71 | 10.0 | 533 1 | DHAX_YEAST |
| 18 | 70.5 | 9.9 | 546 1 | STR_THENA |
| 19 | 70.5 | 9.9 | 670 1 | C21D_HUMAN |
| 20 | 70.5 | 9.9 | 700 1 | CH60_PLAFG |
| 21 | 70.5 | 9.9 | 899 1 | FURS_DROME |
| 22 | 70.5 | 9.9 | 1101 1 | FURS_DROME |
| 23 | 70 | 9.8 | 222 1 | Y175_MERJA |
| 24 | 70 | 9.8 | 948 1 | SECA_STNP7 |
| 25 | 69.5 | 9.8 | 235 1 | SPEC_STNP7 |
| 26 | 69.5 | 9.8 | 611 1 | SNF1_CANGA |
| 27 | 69.5 | 9.8 | 1376 1 | RPOD_ARATH |
| 28 | 69.5 | 9.8 | 2366 1 | TOXB_CLODI |
| 29 | 69 | 9.7 | 132 1 | IDHG_SCHPO |
| 30 | 69 | 9.7 | 430 1 | OSTB_YEAST |
| 31 | 69 | 9.7 | 471 1 | KP19_ARATH |
| 32 | 69 | 9.7 | 660 1 | P750_MYCPN |
| 33 | 69 | 9.7 | 691 1 | PP21_YEAST |

| | | | | | |
|----|------|-----|--------|-------------|--------------------|
| 34 | 69 | 9.7 | 1597 1 | GTF1_STRDO | P11001 streptococ |
| 35 | 68.5 | 9.6 | 482 1 | IFET5_HUMAN | Q13325 homo sapien |
| 36 | 68.5 | 9.6 | 753 1 | NECI_HUMAN | P29120 homo sapien |
| 37 | 68 | 9.6 | 220 1 | GSPE_ERWCH | Q01563 erwina chr |
| 38 | 68 | 9.6 | 379 1 | ACDS_CLOAB | P52042 clostridium |
| 39 | 68 | 9.6 | 352 1 | HMD1_DICDI | P34135 dictyostell |
| 40 | 68 | 9.6 | 719 1 | YP62_CAEEL | Q09437 caenorhabd1 |
| 41 | 68 | 9.6 | 932 1 | SECA_STNP3 | Q05709 synchocyst |
| 42 | 68 | 9.6 | 1267 1 | DHRI_YEAST | Q04217 saccharomyc |
| 43 | 67.5 | 9.5 | 645 1 | SDBE_BACSU | P16396 bacillus su |
| 44 | 67.5 | 9.5 | 775 1 | VP4_ROTIV | P11194 human rotav |
| 45 | 67.5 | 9.5 | 798 1 | ITB1_MOUSE | P09055 mus musculu |

ALIGNMENTS

RESULT 1
ID CALU_MOUSE STANDARD; PRT; 315 AA.
AC Q35887;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CALUMENIN PRECURSOR.
GN CALU;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
RC STRAIN=ICR; TISSUE=Heart.
RX MEDLINE=97364750; PubMed=9218460.
RA Yabe D., Nakamura T., Kanazawa N., Tashiro K., Honjo T.;
RT "Calumenin, a Ca2+-binding protein retained in the endoplasmic
reticulum with a novel carboxyl-terminal sequence, HDEF.";
RJ J. Biol. Chem. 272:18232-18239(1997).
RL
CC -1- FUNCTION: NOT KNOWN. BINDS CALCIUM.
CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.
CC -1- SIMILARITY: BELONGS TO THE RETICULOCALBIN FAMILY.
CC -1- SIMILARITY: CONTAINS 6 EF-HAND CALCIUM-BINDING DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: U01829; AAC53316.1; -
CC MGD: MGI:1097156; Calu.
CC InterPro: IPR002048; -
CC Pfam: PF00036; ehand; 2.
CC PROSITE: PS00018; EF_HAND_4;
CC Calcium-binding; Endoplasmic reticulum; Signal; Repeat.
CC SIGNAL
CC CHAIN
CC FT CA:BIND 20 315
CC FT CA:BIND 81 92
CC FT CA:BIND 117 128
CC FT CA:BIND 164 175
CC FT CA:BIND 201 212
CC FT CA:BIND 242 253
CC FT CA:BIND 278 289
CC FT CARBOHYD 131 131
CC FT SITE 312 315
CC SEQUENCE 315 AA; 37063 MM; 742361814171E273 CRC64;
Query Match 11.18; Score 79; DB 1; Length 315;
Best Local Similarity 22.18; Pred. No. 4.3;
Matches 42; Conservative 32; Mismatches 46; Indels 70; Gaps 11;

```

QY 1 IVSLSLC-AVALAQ-----HRSQENKDNRRVSYDQSSQSKSENL 41
DB 7 IMCLSLCTAVALASPEKEDKRVHEPQSDKVNDAQNFYDHAFL-GAEKAKSPQLT 65
QY 42 PDVSOKEGIAQAEIVYKTD--QGYT-----SHGHYHYNG-KVPYD 83
DB 66 PEERKEKIG---KIVSKIDDDKGFVTVDELKGMFAKRWIHEVEKQMGHDJNED 121
QY 84 ALFSEE-----LAKMD-----NYOL-----KDAIVNEVGGYLIKY 116
DB 122 GLVSWSEYKNAATGYVLDPPDDGPFNYKQMAVDERFRFADKDGDL-ATKEEFAFL 160
QY 117 DGKYYVYKLD 126
DB 181 HPEEYDMKD 190

RESULT 2
PPID_HAEIN STANDARD; PRT; 622 AA.
ID PPID_HAEIN STANDARD; PRT; 622 AA.
AC P44092;
DE 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
PE 01-OCT-2000 (Rel. 40, Last annotation update)
PR PEPTIDYL-PROLYL CIS-TRANS ISOMERASE D (EC 5.2.1.8) (PIPIASE D)
(ROTAMASE D).
PPID OR HI1004.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX Haemophilus.
RN NCBI_Taxid=727;
RP SEQUENCE FROM N.A.
RC STRAIN=RD / K20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7342800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kierulff A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shrivley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrman J.L., Geoghagen N.S.M.,
RA Grehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT *whole-genome random sequencing and assembly of Haemophilus
RL Influenzae Rd. Science 269:496-512(1995).
RN [2]
RP CONCEPTUAL TRANSLATION.
RA Balroch A.;
RL Unpublished observations (MAY-1998).
CC -1- FUNCTION: PIPASES ACCELERATE THE FOLDING OF PROTEINS. SEEMS TO BE
INVOLVED IN THE FOLDING OF OUTER MEMBRANE PROTEINS (BY
SIMILARITY).
CC -1- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
PEPTIDE BONDS IN OLIGOPEPTIDES.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE PPIC/PARVULIN FAMILY OF ROTAMASES.
STRONG, TO E.COLI PPID.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS
INTRODUCED IN POSITION 43 TO PRODUCE THIS ORF.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
in the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
CC CC
DR EMBL: U37781; AAC22665.1; ALT_FRAME.
DR TIGR: HI1004;

```

```

DR InterPro: IPR000297;
DR Pfam: PF00639; Rotamase; 1.
DR PROSITE: PS01096; PPIC_PPIASE_1; 1.
DR PROSITE: PS01098; PPIC_PPIASE_2; 1.
KW Isomerase; Rotamase; Transmembrane; Periplasmic.
FT DOMAIN 1 16 CTROPPLASMIC (POTENTIAL).
FT TRANSMEM 17 37 POTENTIAL.
FT DOMAIN 38 622 PERIPLASMIC (POTENTIAL).
FT DOMAIN 270 356 PPIC-LIKE.
SQ SEQUENCE 622 AA; 69590 MW; EES90KA5381C4B4 CRC64;

```

Query Match 10.8%; Score 76.5; DB 1; Length 622;
 Best Local Similarity 21.0%; Pred. No. 16;
 Matches 37; Conservative 29; Mismatches 67; Indels 43; Gaps 6;

```

QY 1 IVSLSLC-AVALAQ-----KDNRRVSYDQSSQSKSENLTPDVSOKEGIAQAEI 56
DB 196 IATLSLADEMAKQSVSDDEIKTYEANKQSFVQPEQVKYIDLSADNISRNLQYDVEI 235
QY 57 VIKITQO--GYTSHGDIHYNGKVPYDALFSEELMDPNY---QLKDAIVNEVKG 110
DB 256 AQYQDKAKQFMTQHLAHIOFANEQ--DAKVAIEELQGANADYAKAKSLDKISGNG 312
QY 111 GYI-----IKYDKRYVYL---KDAHADNVRXK 136
DB 313 GDLGWVNEENLPRAFEDAAALQVGQYSPINVDGNVHYLVQERRAKQSLQENKAO 368

```

```

RESULT 3
BGAL_THERET STANDARD; PRT; 743 AA.
ID BGAL_THERET STANDARD; PRT; 743 AA.
AC P77989;
DE 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
PE 15-DEC-1998 (Rel. 37, Last annotation update)
DE BETA-GALACTOSIDASE (EC 3.2.1.23) (LACYASE).
GN LACZ OR LACA.
OS Thermomanaerobacter ethanolicus (Clostridium thermohydrosulfuricum).
OC Bacteria; Firmicutes; Bacilli; Clostridium group;
OX Thermomanaerobacter group; Thermomanaerobacter.
RN NCBI_Taxid=1757;
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33223 / 39E;
RA Zverlov V.;
RT Submitted (FEB-1998) to the EMBL/GenBank/DBS databases.
RL -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.
CC CC

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- in the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC CC
CC EMBL: Y08557; CAA69850.1;
DR InterPro: IPR001649;
DR Pfam: PF00703; Glyco_hydro_2; 1.
DR PROSITE: PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE: PS00608; GLYCOSYL_HYDROL_F2_2; 1.
KW Hydrolase; Glycosidase.
FT ACT_SITE 388 388 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 453 453 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 743 AA; 85796 MW; FE011FE51JESIDFC CRC64;

```

Query Match 10.5%; Score 74.5; DB 1; Length 743;
 Best Local Similarity 27.8%; Pred. No. 31;
 Matches 25; Conservative 14; Mismatches 30; Indels 21; Gaps 4;

QY 44 QVSKKEGI-----GAGVIVIKINDGVTSHGDHYHYNGKVPYDALFSEELM 92
 DB 198 EVINKEEVGVSSVVTDIKEKQKEIVQIKNDNPLTWPHDPHYLVNSV---KLIAEELL 254
 QY 93 KDPYOLKADIVNEVGGYIIKVDKRYV 122
 DB 255 --DNVTFKRG-----IKGVYERDDGKFFYI 277
 RESULT 4
 UME6 YEAST STANDARD; PRT; 836 AA.
 ID UME6 YEAST STANDARD; PRT; 836 AA.
 P39001;
 01-FEB-1995 (Rel. 31, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-NOV-1997 (Rel. 35, Last annotation update)
 DE TRANSCRIPTIONAL REGULATOR UME6 (NEGATIVE TRANSCRIPTIONAL REGULATOR OF
 TIME2).
 GN UME6 OR CAR80 OR CAR81 OR NIM2 OR YDR207C OR YB8142.04C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycaceae.
 NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S1278B;
 RA Smart W.C., Park H.-D., Cooper T.G.;
 RL Submitted (XXX-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=95011581; PubMed=7926768;
 RA Strich R., Surosky R.T., Steber C., Messenguy F., Dubois E.,
 RA Easton Esposito R.;
 RT "UME6 is a key regulator of nitrogen repression and meiotic
 development.";
 RT Genes Dev. 8:796-810(1994).
 RL [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-GNF88;
 RA Kumeno A.;
 RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Oliver K., Shore L., Harris D., Barrell B.G., Rajandream M.A.,
 RA Welsh S.V.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP CHARACTERIZATION.
 RC MEDLINE=92253412; PubMed=1579492;
 RA Park H.-D., Luche R.M., Cooper T.G.;
 RA "The yeast UME6 gene product is required for transcriptional
 repression mediated by the CAR1 URS1 repressor binding site.";
 RT Nucleic Acids Res. 20:1909-1915(1992).
 RL [6]
 RP CHARACTERIZATION.
 RC MEDLINE=96100650; PubMed=8528081;
 RA Anderson S.F., Steber C.M., Easton Esposito R., Coleman J.E.;
 RA "UME6, a negative regulator of meiosis in Saccharomyces cerevisiae,
 RT contains a C-terminal Zn2Cys6 binuclear cluster that binds the URS1
 RT DNA sequence in a zinc-dependent manner.";
 RT Protein Sci. 4:1832-1843(1995).
 RL [7]
 RP CHARACTERIZATION.
 RC MEDLINE=9618868; PubMed=8614637;
 RA Jackson J.C., Lopes J.M.;
 RA "The yeast UME6 gene is required for both negative and positive
 RT transcriptional regulation of phospholipid biosynthetic gene
 RT expression.";
 RT Nucleic Acids Res. 24:1322-1329(1996).
 CC -1- FUNCTION: BINDS TO THE URS1 SITE (5'-AGCGCGCA-3') AND NEGATIVELY
 CC REGULATES THE EXPRESSION OF MANY GENES INCLUDING CAR1 (ARINASE),

CC SEVERAL REQUIRED FOR SPOULATION, MATING, TYPE SWITCHING, INOSITOL
 CC METABOLISM, AND OXIDATIVE CARBON METABOLISM. HAS BOTH A POSITIVE
 CC AND NEGATIVE ROLE IN REGULATING PHOSPHOLIPID BIOSYNTHESIS.
 CC -1- SUBUNIT: INTERACTS WITH IME1.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
 CC CLUSTER DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.slb.ch).
 CC
 CC EMBL; J2186; AAA34471.1;
 CC EMBL; L24539; AAC14472.1;
 CC EMBL; D23663; BAA04890.1;
 CC EMBL; Z68194; CAA92346.1;
 CC HSSP; P12351; 1PYC.
 CC TRANSFAC; T01247;
 CC SGD; S0002615; UME6.
 CC InterPro; IPR001138;
 CC Pfam; PF00172; Zn_C1us; 1.
 CC PROSITE; PS00463; ZN2_C1us; 1.
 CC PROSITE; PS00463; ZN2_C1us_FUNGAL_2; 1.
 CC Transcription regulation; Repressor; activator; DNA-binding;
 CC Nuclear protein; Zinc; Metal-binding.
 CC FT DNA_BIND 771 798 ZN(2)-CYS(6), FUNGAL-TYPE.
 CC FT CONFLICT 101 101 V -> G (IN REF. 3).
 CC FT CONFLICT 363 363 I -> V (IN REF. 1).
 CC FT CONFLICT 443 443 N -> T (IN REF. 1).
 CC FT CONFLICT 465 465 G -> D (IN REF. 1).
 CC FT CONFLICT 465 465 G -> D (IN REF. 1).
 CC SQ SEQUENCE 836 AA; 91123 MW; 0DDA0A0B4A157182 CRC64;
 Query Match 10.5%; Score 74.5; DB 1; Length 836;
 Best Local Similarity 21.6%; Pred. No. 35;
 Matches 29; Conservative 23; Mismatches 43; Indels 39; Gaps 5;
 QY 11 INCHRSQENDNRKVSIVDSQSSQSKSE-----NLTPOVSO-REGIOAEQIVL 58
 DB 275 LSRNNSTNNNDNNSIOSDSRESNNNEIGYLRGTRKRGSPSNDQVGHVHDDQCAV 334
 QY 59 KINDGVYSHGDHYHYNGKVPYDALFSEELMKDPNOLKADIVNEVGGYIIKVDG 118
 DB 335 -----GVAPRNFYFNKDRREITDPNVXIDE-----NESKINISFWLNS 371
 QY 119 KYTVYIKDAHADN 132
 DB 372 KY-----RDEAVSLN 381
 RESULT 5
 PCEL_HUMAN STANDARD; PRT; 710 AA.
 ID PCEL_HUMAN STANDARD; PRT; 710 AA.
 AC P48147;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE PROLYL ENDOPEPTIDASE (EC 3.4.21.26) (PROT-PROLINE CLEAVING ENZYME)
 DE (PE)
 GN PREP OR PEP.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Blood;
 RX MEDLINE=95047504; PubMed=7959018;
 RA Vanhoof G., Goossens F., Hendriks L., de Meester I., Hendriks D.,

RA Vriend G., van Broekhoven C., Scharpe S.,
 RT "Cloning and sequence analysis of the gene encoding human lymphocyte
 RT prollyl endopeptidase."
 RL Gene 149:363-366(1994).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-94375419; PubMed-8089089;
 RA Shiraawa Y., Osawa T., Hiraishima A.,
 RT "Molecular cloning and characterization of prollyl endopeptidase from
 RT human T cells."
 RL J. Biochem. 115:724-729(1994).
 CC -1- FUNCTION: CLEAVES PEPTIDE BONDS ON THE C-TERMINAL SIDE OF PROLYL
 CC RESIDUES WITHIN PEPTIDES THAT ARE UP TO APPROXIMATELY 30 AMINO
 CC ACIDS LONG.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF PRO-|-XAA >> ALA-|-XAA IN
 CC OLIGOPEPTIDES.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9A; ALSO KNOWN AS THE
 CC PROLYL OLIGOPEPTIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X74496; CA52605.1; -
 CC EMBL: D21102; BAA04661.1; -
 CC MEROPS: S09.001; -
 CC MIM: 600400; -
 CC InterPro: IPR001375; -
 CC InterPro: IPR002470; -
 CC InterPro: IPR002471; -
 CC Pfam: PF00326; Peptidase_S9; 1.
 CC PRINTS: PR00862; PROLIGOPASE.
 CC PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
 CC KMW Hydrolyase; Serine protease.
 CC FT ACT_SITE 554 554 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC FT ACT_SITE 641 641 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC FT ACT_SITE 680 680 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC FT CONFLICT 16 16 V -> I (IN REF. 2).
 CC FT CONFLICT 245 245 T -> C (IN REF. 2).
 CC FT CONFLICT 298 298 T -> A (IN REF. 2).
 CC FT CONFLICT 319 319 W -> R (IN REF. 2).
 CC FT CONFLICT 440 440 I -> L (IN REF. 2).
 CC FT CONFLICT 459 459 G -> S (IN REF. 2).
 CC FT CONFLICT 459 459 V -> I (IN REF. 2).
 CC FT CONFLICT 706 706
 CC SQ SEQUENCE 710 AA; 80763 MW; 241D0F2D761A8DD2 CRC64;

Query Match 10.4%; Score 74; DB 1; Length 710;
 Best Local Similarity 25.8%; Pred. No. 32;
 Matches 32; Conservative 14; Mismatches 46; Indels 32; Gaps 6;

DB 42 PDVVSQREGIOA-----EQLVIR-----ITDGYVTSR--GDHYHY----76
 DB 34 PDSEQFAVEAQNKIIVPLEQCPINGLTKERNTLEYDIPKSCHEKRRFFYFNTG 93
 DB 77 --NGKVPY--DALFSEELMKDPNYQLKADYINVEKGYITKDKYVYLLADAHADN 132
 DB 94 LQNGRVLYVODSLGEARVFLDPNLLSDGTVLR--GFAFSEDEYFAVGLSAGSDW 150
 DB 133 VRTK 136
 DB 151 VTRK 154
 RESULT 6
 PCRE_PIG ID PCRE_PIG STANDARD; PRT; 710 AA.
 AC P23687;

DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE PROLYL ENDOPEPTIDASE (EC 3.4.21.26) (POST-PROLINE CLEAVING ENZYME)
 DE (PE)
 GN PREP.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 OX NCBI_TaxID:9823;
 RN [1]
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ACTIVE SITE SER-554.
 RC TISSUE-Brain;
 RX MEDLINE-91152034; PubMed-1900195;
 RA Rennex D., Hemmings B.A., Hofsteenge J., Stone S.R.;
 RT Identification of porcine brain prollyl endopeptidase and
 RT identification of the active-site seryl residue.
 RL Biochemistry 30:2195-2203(1991).
 [2]
 RP ACTIVE SITE HIS-680.
 RX MEDLINE-91291146; PubMed-2064618;
 RA Stone S.R., Rennex D., Wikstrom P., Shaw E., Hofsteenge J.;
 RT Inactivation of prollyl endopeptidase by a peptidylchloromethane.
 RT Kinetics of inactivation and identification of sites of
 RT modification.
 RL Biochem. J. 276:837-840(1991).
 CC -1- FUNCTION: CLEAVES PEPTIDE BONDS ON THE C-TERMINAL SIDE OF PROLYL
 CC RESIDUES WITHIN PEPTIDES THAT ARE UP TO APPROXIMATELY 30 AMINO
 CC ACIDS LONG.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF PRO-|-XAA >> ALA-|-XAA IN
 CC OLIGOPEPTIDES.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- TISSUE SPECIFICITY: IN ALL TISSUES.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9A; ALSO KNOWN AS THE
 CC PROLYL OLIGOPEPTIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M64227; AAA31110.1; -
 CC PIR: A37942; A37942.
 CC MEROPS: S09.001; -
 CC InterPro: IPR001375; -
 CC InterPro: IPR002470; -
 CC InterPro: IPR002471; -
 CC Pfam: PF00326; Peptidase_S9; 1.
 CC PRINTS: PR00862; PROLIGOPASE.
 CC PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
 CC KMW Hydrolyase; Serine protease.
 CC FT ACT_SITE 554 554 CHARGE RELAY SYSTEM.
 CC FT ACT_SITE 641 641 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC FT ACT_SITE 680 680 CHARGE RELAY SYSTEM.
 CC FT VARIANT 29 29 A -> H.
 CC SQ SEQUENCE 710 AA; 80769 MW; 70286A86238D72C0 CRC64;

Query Match 10.4%; Score 74; DB 1; Length 710;
 Best Local Similarity 25.8%; Pred. No. 32;
 Matches 32; Conservative 14; Mismatches 46; Indels 32; Gaps 6;

DB 42 PDVVSQREGIOA-----EQLVIR-----ITDGYVTSR--GDHYHY----76
 DB 34 PDSEQFAVEAQNKIIVPLEQCPINGLTKERNTLEYDIPKSCHEKRRFFYFNTG 93
 DB 77 --NGKVPY--DALFSEELMKDPNYQLKADYINVEKGYITKDKYVYLLADAHADN 132
 DB 94 LQNGRVLYVODSLGEARVFLDPNLLSDGTVLR--GFAFSEDEYFAVGLSAGSDW 150

OY 133 VTRK 136
DB 151 VTRK 154

RESULT 7

PRML_BRARE STANDARD; PRT; 713 AA.

AC Q9W735;

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

PROMININ-LIKE PROTEIN (FRAGMENT)

Brachydanio rerio (Zebrafish) (Zebra danio).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Rasbora; Danio.

NCBI_Taxid=7955;

[1]

SEQUENCE FROM N.A.

TISSUE=Liver;

Corbell D., Treichel J., Roper K., Brand M., Hutter W.B.;

"Sequence analysis of a zebrafish ortholog of mouse prominin and human

Actin antigen."

Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

-1- SIMILARITY: BELONGS TO THE PROMININ FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL, AF160970; AAD4341.1;

InterPro: IPR001522;

Transmembrane; Glycoprotein.

TRANSMEM 50 70

TRANSMEM 106 126

TRANSMEM 153 173

TRANSMEM 439 459

TRANSMEM 483 503

TRANSMEM 178 178

CARBOHYD 268 268

CARBOHYD 286 286

CARBOHYD 327 327

CARBOHYD 388 388

CARBOHYD 404 404

CARBOHYD 576 576

CARBOHYD 582 582

CARBOHYD 617 617

CARBOHYD 693 693

NON_TER 713 713

SEQUENCE 713 AA; 78990 MW; 8C7EBDDE3039BCD CRC64;

Query Match 10.3%; Score 73.5; DB 1; Length 713;

Best Local Similarity 35.5%; Pred. No. 36;

Matches 35; Conservative 19; Mismatches 66; Indels 17; Gaps 5;

6 LCAVALNDRSQRNDNRVYVDGSSQSKS-----ENTLPPDVS---OKEGIGAEQIVY 58

170 LCAVALNDRSQRNDNRVYVDGSSQSKS-----VKSNLKDLTPANQPAQIDYLSRGVKEQVLA 224

59 KITDGGVTSNGDHYHYNG---KVPYDALFSEELKMDPYQAKADIVNEVGGYIIR 115

225 DLENNGVIL--GGRHIEELGKEVVPALDALTSMGTMRDRLDALENNVSLLETLOEGTVK 282

116 VDGKYYVYLKDAAHND 132

DB 283 LQANLTVYRNSLRNALN 299

RESULT 8

SENAPLCA STANDARD; PRT; 113 AA;

AC P29233;

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

SENORIN A PRECURSOR.

PSCL

OS Aplysia californica (California sea hare).

Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidae;

Aplysiidae; Aplysia.

NCBI_Taxid=6500;

[1]

SEQUENCE FROM N.A., AND SEQUENCE OF 33-54.

TISSUE=Pleural sensory cells;

RC MEDLINE-91227915; PubMed-1840700;

RT Brunet J.-F., Shapiro E., Foster S.A., Kandel E.R., Iino Y.;

"Identification of a peptide specific for Aplysia sensory neurons by

PCR-based differential screening."

Science 252:856-859(1991).

-1- FUNCTION: MAY FUNCTION AS AN INHIBITORY COTRANSMITTER ACTING IN

CONJUNCTION WITH THE FAST EXCITATORY TRANSMITTER RELEASED BY

SENSORY NEURONS. THE PEPTIDE SELECTIVELY INHIBITS CERTAIN

POSTSYNAPTIC CELLS PROBABLY BY MEANS OF SENORIN A RELEASE.

-1- SUBCELLULAR LOCATION: THROUGHOUT THE NEURONAL CELLS (CELL BODY,

AXON & PRESYNAPTIC TERMINALS).

TISSUE SPECIFICITY: SEEMS TO BE SPECIFIC TO THE MECHANOSENSORY

NEURONS OF THE CENTRAL NERVOUS SYSTEM.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL, X56770; CAA0089.1;

DR PIR: S23653; S23653.

DR HSP: P02586; 2TN4.

KW SIGNAL; Neutrone; Amidation; Cleavage on pair of basic residues.

FT PEPTIDE 33 54

FT MOD RES 54 54

SEQUENCE 113 AA; 12711 MW; 74350F5154B49E1A CRC64;

Query Match 10.3%; Score 73; DB 1; Length 113;

Best Local Similarity 32.6%; Pred. No. 42;

Matches 29; Conservative 14; Mismatches 24; Indels 22; Gaps 5;

OY 1 TVSLSLCAVALNDRSQRNDNRVYVDGSSQSKS-----KDNKRYSTYDGSQSSQSKS-----NLTPPVYSQKE 49

DB 20 IYCALQAVANANRNSNNVPRPRARYRGVYGFGRSSSEYTSNLIINLSQVSOE 79

OY 50 GION-----VIRI---TPDGYVT 67

DB 80 ELRAILKEQPIILDEYVKKIIRNDGDIIT 108

RESULT 9

Y943 META STANDARD; PRT; 417 AA.

AC Q88353;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE HYPOTHETICAL PROTEIN MJ0943.

GN M0943.
 OS Methanococcus jannaschii.
 CC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 CC Methanococcus.
 RX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RA MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kere J., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness N.S., Weinstock K.G., Merrick J.M., Nguyen A.,
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrman J.L., Nguyen A.,
 RA Uitterlinden T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RT Science 273:1058-1073(1996).
 RL -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U67538; AAB98953.1;
 DR TIGR: M0943;
 DR Hypothetical protein; Transmembrane.
 FT TRANSMEM 10 30 POTENTIAL.
 FT TRANSMEM 148 168
 FT DOMAIN 399 415 PRO-RICH.
 SQ SEQUENCE 417 AA; 46355 MW; 6DF8D82E1AABD90 CRC64;

Query Match 10.3%; Score 73; DB 1; Length 417;
 Best Local Similarity 22.9%; Pred. No. 21;
 Matches 33; Conservative 23; Mismatches 43; Indels 42; Gaps 6;

QY 16 SQENKDNRRSYDYDSSQSSQSENLPPDOVSQKEIOAEQIVIRITDQGYTSGHGR--- 72
 DB 52 NQENQNNNNENKESQSTQISQENKEIKQENHPLQ-----SQNQNQNGNNNEE 105
 QY 73 -----YHT-----NGKVPYDALFSEELMKDPYQLKADIVN 106
 DB 106 NENAMTNVGESEVYNNNEPANNYIEITPDGTRP-DKI-EEQWL-----YIKVIDPIV 158
 QY 107 EVKGGYIIIVDGRYYVYKLD 126
 DB 159 GGLAGIDIVDGNITGLTD 178

RESULT 10
 E6_GOSHI
 ID E6_GOSHI STANDARD; PRT; 238 AA.
 AC 001197;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUN-1998 (Rel. 36, Last annotation update)
 DE PROTEIN E6.
 GN E6.
 OS Gossypium hirsutum (Upland cotton).
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 CC Malvales; Malvaceae; Gossypium.
 OX NCBI_TaxID=3635;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COKER 312; TISSUE-Fiber;

RX MEDLINE=92335179; PubMed=1631059;
 RA John M.E., Crow L.J.
 RT "Gene expression in cotton (Gossypium hirsutum L.) fiber: cloning of
 RT the mRNA.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:5769-5773(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COKER 312;
 RX MEDLINE=96178868; PubMed=8616253;
 RA John M.E.;
 RT "Structural characterization of genes corresponding to cotton fiber
 RT mRNA, E6: reduced E6 protein in transgenic plants by antisense
 RT gene.";
 RL Plant Mol. Biol. 30:297-306(1996).
 CC -1- TISSUE SPECIFICITY: IT IS PREDOMINANTLY EXPRESSED IN FIBER CELLS.
 CC -1- DEVELOPMENTAL STAGE: CONCENTRATION OF E6 IS HIGHEST DURING THE
 CC LATE PRIMARY CELL WALL AND EARLY CELL WALL SYNTHESIS STAGES.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-5, MET-10 OR MET-21 IS
 CC THE INITIATOR.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M92051; AAA33055.1;
 DR EMBL: M92051; AAA33056.1; ALT_INIT.
 DR EMBL: U30505; AAB03079.1;
 DR PIR: A46130; A46130.
 DR Cell wall.
 SQ SEQUENCE 238 AA; 28226 MW; 7162101A9A91AFB CRC64;

Query Match 10.1%; Score 72; DB 1; Length 238;
 Best Local Similarity 26.4%; Pred. No. 13;
 Matches 28; Conservative 12; Mismatches 40; Indels 26; Gaps 5;

QY 16 SQENKDNRRSYDYDSSQSSQSENLPPDOVSQKEIOAEQIVIRITDQGYTSGHGRH 75
 DB 116 SSNKKDT--YYKNKNAYESKQOOL-----GEKI--FEKGGSTENQNNNT 158
 QY 76 YNKKVPYDALFSEELMKDPYQLKADIVNEVKGGIIVDGRYY 121
 DB 159 YNNGNGTNN--GEKQMSDTRT-----LENGKYYIVKASENNY 195

RESULT 11
 EAS_DROME
 ID EAS_DROME STANDARD; PRT; 517 AA.
 AC P54352; O9YX17;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ETHANOLAMINE KINASE (EC 2.7.1.82) (E6) (EASILY SHOCKED PROTEIN).
 GN EAS OR CG3525.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE-Embryo;
 RX MEDLINE=95007778; PubMed=7923374;
 RA Pavlidis P., Ramaswami M., Tanouye M.A.;
 RT "The Drosophila easily shocked gene: a mutation in a phospholipid
 RT synthetic pathway causes seizure, neuronal failure, and paralysis.";
 RL Cell 79:23-33(1994).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).

RC STRAIN-BERKELEY; PubMed-10731132;
RA MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton R.G., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Branton R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Bokoyva D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Butts J.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dev I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Stryckas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Wetsenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*.";
RA Science 287:2185-2195(2000).
CC -1- CATALYTIC ACTIVITY: ADP + ETHANOLAMINE - ADP + O-
CC -1- PHOSPHOETHANOLAMINE.
CC -1- PATHWAY: PHOSPHATIDYLETHANOLAMINE SYNTHESIS.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOPROPS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- MISCELLANEOUS: MUTATIONS IN EAS CAUSES SEIZURE, NEURONAL FAILURE,
CC AND PARALYSIS. THIS IS DUE TO AN EXCITABILITY DEFECT CAUSED BY
CC ALTERED MEMBRANE PHOSPHOLIPID COMPOSITION.
CC -1- SIMILARITY: BELONGS TO THE CHOLINE/ETHANOLAMINE KINASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; L35603; AAC37209.1;
CC EMBL; L35604; AAC37210.1;
CC EMBL; AE003501; AAF48574.1;
CC FLYBASE; FBgn0000536; eas.
CC InterPro; IPR002573;
CC DR P1fam; PF01633; Choline_kinase; 1.
CC K0 Transferrase; Kinase; Alternative splicing.
CC ACT SITE 374 374 BY SIMILARITY.
CC FT VARSPLIC 130 151 MISSING (IN SHORT ISOFORM).
CC FT SEQUENCE 517 AA; 59208 MW; BA35636B07D6786 CRC64;
CC SQ

Query Match 10.1%; Score 72; DB 1; Length 517;
Best Local Similarity 22.9%; Pred. No. 33;
Matches 39; Conservative 16; Mismatches 43; Indels 72; Gaps 7;

QY 3 SLICAVANLMOH-----RSQENKNNRVSYDSSQSSQSEN-----LTPDQ 44
Db 27 SLISVROVNOQTSLSSQNOVONLNSNSNSPNSGSENNENBONSNDIRAKPED 86
QY 45 VSQKEGIAQEDIVIKITIDGVTYSHGHYHYNGKVPDALFSE-----LAKD 94
Db 87 KSRKEAI-----VPEVPPFEVADVIQAKELKAYR 118
QY 95 PNYQLKDD-----TVNEVKGIIKVDGKYVYLKDAHADNVKTK 136
Db 119 PTWLSHVEKIRVPPQIEDRVSGP---KCDGD-----DDASFDTGTYNK 160
RESULT 12
SLPH_BACBR STANDARD; PRT; 1116 AA.
ID SLPH_BACBR
AC P38538;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE SURFACE LAYER PROTEIN PRECURSOR (HEXAGONAL WALL PROTEIN) (HWP).
OS Bacillus brevis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Brevibacillus.
OX NCBI_TaxID-1393;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 54-68.
RC STRAIN-HPD31.
RX MEDLINE-90170842; PubMed-2307650;
RA Edisu S., Tsuboi A., Takagi H., Naruse Y., Yamagata H., Tsukagoshi N.,
RA Ueda S.;
RT "Conserved structures of cell wall protein genes among
RT protein-producing Bacillus brevis strains.";
RL J. Bacteriol. 172:1312-1320(1990).
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC HEXAGONAL S-LAYER.
CC -----
CC -1- SIMILARITY: CONTAINS 3 S-LAYER HOMOLOGY (SLH) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; D90050; BA14103.1;
CC EMBL; A35129; A35129.
CC InterPro; IPR001119;
CC DR P1fam; PF00395; SLH; 2.
CC DR PROSITE; PS01072; SLH_DOMAIN; 2.
CC K0 Signal; Cell wall; S-layer; Repeat.
CC FT SIGNAL 5 53
CC FT CHAIN 54 1116 SURFACE LAYER PROTEIN.
CC FT DOMAIN 57 120 SLH 1.
CC FT DOMAIN 121 171 SLH 2.
CC FT DOMAIN 172 231 SLH 3.
CC FT SEQUENCE 1116 AA; 123397 MW; 86D583D7AC72546F CRC64;
CC SQ

Query Match 10.1%; Score 72; DB 1; Length 1116;
Best Local Similarity 28.1%; Pred. No. 84;
Matches 34; Conservative 23; Mismatches 46; Indels 18; Gaps 8;
QY 22 NNRVSYDVSQSSQSENLTPDQVSQKEGIAQEDIVIKITIDGVTYSHGHYHYNGRY- 80
Db 872 NN--NHIDNTAS--KSATLMPEDERQKGT--DKYVARVDE--VDGWTISLNTVADGKTQ 923
QY 81 PYDALFSEELMKDPNOLKADIVNEVKGIIYI---KYDGKTYVYLKDAHADNVKTK 135
Db 924 KYTKASTAFI--DYYDLGLIGDGVDE--GDYIWDISADIDGTFEDVYLVVSSDDEIRT 979

OY 136 K 136
DB 980 Q 980

RESULT 13

GTFF2_STRDO STANDARD; PRT; 1592 AA.
ID GTFF2_STRDO
AC P27470;
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=1317;
RX MEDLINE-91123227; PubMed-1704006;
RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,
Kagawa H.;
RA "peptide sequences for sucrose splitting and glucan binding within
Streptococcus sobrinus glucosyltransferase (water-insoluble glucan
synthetase)".
RA J. Bacteriol. 173:989-996(1991).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE), GTF-SI SYNTHESIZES BOTH
FORMS OF GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
BINDING PROTEIN FROM S. MOTANS.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90213; BAA14241.1;
DR PIR; A38175; A38175.
DR HSP; P00695; 2HEE.
InterPro: IPR002479;
Pfam: PF01473; CW-binding_1; 16.
KW Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 38 POTENTIAL.
FT CHAIN 39 1592 GLUCOSYLTRANSFERASE-1.
FT DOMAIN 39 1044 CATALYTIC (APPROXIMATE).
FT DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).
FT REPEAT 1093 1592 6.5 X TANDEM REPEATS.
FT REPEAT 1158 1207 1.
FT REPEAT 1222 1272 2.
FT REPEAT 1287 1337 3.
FT REPEAT 1402 1451 4.
FT REPEAT 1514 1563 5.
FT REPEAT 1577 1592 6.
SO SEQUENCE 1592 AA; 176167 MW; BCAA66D079351ECF CRC64;
7 (INCOMPLETE).
Query Match 10.18; Score 72; DB 1; Length 1592;
RT

Best Local Similarity 23.28; Pred. No. 1.3e+02;
Matches 36; Conservative 18; Mismatches 45; Indels 56; Gaps 7;

OY 9 YALNHSOEN-----KONNR---VSYYDGSOSKSENFPTDYOSQRECI 51
DB 1137 YYANDGRYENENGYOQFGNDWRYFKDGNMNAVGLTYDGNV-----QFDDKGV 1185
OY 52 QAEIVIKITPDGYVTSQHD-----YHYNGK---VPYDALSEELIM 92
DB 1186 QAKRKII-VTDGKVRFRFDQNGNAYNTFTADTGKHYIYIGKGVAVTGQVYGRKLY 1244
OY 93 KDPYQLKADIVNEVGKYYIKYDGRYYVYLKDA 127
DB 1245 FEANGQ-----QKGDFTVSDGKLYFYDVS 1271

RESULT 14

NECI_MOUSE STANDARD; PRT; 753 AA.
ID NECI_MOUSE
AC P21662; P22546;
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-AUG-1991 (Rel. 19, Last annotation update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEUROENDOCRINE CONVERTASE 1 PRECURSOR (EC 3.4.21.93) (NEC 1) (PC1)
DE (PROHORMONE CONVERTASE 1) (PROPEPTIDE PROCESSING PROTEIN)
DE PCSK1 OR NEC1 OR NEC-1 OR ATX-1.
OS Mus musculus (Mouse) and Mus cookii.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090; 10098;
RX MEDLINE-91319778; PubMed-1862107;
RA Korner J., Chun J., Harter D., Axel R.;
RT "Isolation and functional expression of a mammalian prohormone
processing enzyme, murine prohormone convertase 1.";
RT Proc. Natl. Acad. Sci. U.S.A. 88:6834-6838(1991).
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-MOUSE;
RX MEDLINE-91203919; PubMed-2017186;
RA Seidah N.G., Marcinkiewicz M., Benjannet S., Gaspar L., Beaulieu G.,
Maret M.-G., Lazure C., Molloy M., Chretien M.;
RT "Cloning and primary sequence of a mouse candidate prohormone
convertase PC1 homologous to PC2, Furin, and Kex2: distinct
chromosomal localization and messenger RNA distribution in brain and
pituitary compared to PC2.";
RT Mol. Endocrinol. 5:111-122(1991).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES-MOUSE;
RX MEDLINE-9110525; PubMed-1988934;
RA Smeekens S.P., Avruch A.S., Lamenola J., Chan S.J., Steiner D.F.;
RT "Identification of a cDNA encoding a second putative prohormone
convertase related to PC2 in R120 cells and 13lets of Langerhans.";
RT Proc. Natl. Acad. Sci. U.S.A. 88:340-344(1991).
RN [5]
RP SEQUENCE OF 214-478 FROM N.A.
RC SPECIES-MOUSE; TISSUE-Pituitary;
RX MEDLINE-91000356; PubMed-2169760;
RA Seidah N.G., Gaspar L., Mion P., Marcinkiewicz M., Molloy M.,
Chretien M.;
RT "cDNA sequence of two distinct pituitary proteins homologous to Kex2
RT

RT and furin gene products: tissue-specific mRNAs encoding candidates
 RT for pro-hormone processing proteolases.";
 RL DNA Cell Biol. 9:415-424(1990).
 RN ERRATUM.
 RC SPECIES-MOUSE;
 RX MEDLINE-91090850; PubMed-2264933;
 RA Seidah N.G., Gaspar L., Mion P., Marcinkiewicz M., Milikay M.,
 Chretien M.;
 RL DNA Cell Biol. 9:789-798(1990).
 CC -1- FUNCTION: INVOLVED IN THE PROCESSING OF HORMONE AND OTHER PROTEIN
 CC PRECURSORS AT SITES COMPRISED OF PAIRS OF BASIC AMINO ACID
 CC RESIDUES. SUBSTRATES INCLUDE POMC, RENIN, ENKEPHALIN, DYNORPHIN,
 CC SOMATOSTATIN AND INSULIN.
 CC -1- CATALYTIC ACTIVITY: RELEASE OF PROTEIN HORMONES, NEUROPEPTIDES AND
 CC RENIN FROM THEIR PRECURSORS, GENERALLY BY CLEAVAGE OF -LIS-ARG-1-
 CC BONDS.
 CC -1- COFACTOR: CALCIUM DEPENDENT.
 CC -1- SUBCELLULAR LOCATION: LOCALIZED IN THE SECRETION GRANULES.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY. HIGH SIMILARITY WITH OTHER FURIN-LIKE ENZYMES.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC
 CC EMBL: M58507; AAA39896.1; -
 CC EMBL: X57088; CAA40368.1; ALT_SEQ.
 CC EMBL: M58588; AAA39894.1; -
 CC EMBL: M58568; AAA39375.1; ALT_SEQ.
 CC EMBL: M69196; AAA39732.1; -
 CC PIR: JX0171; KXMS01.
 CC PIR: A37951; A37951.
 CC MEROPS: S08.072; -
 CC MD: MGI:97511; PcsK1.
 CC InterPro: IPR00209; -
 CC InterPro: IPR002884; -
 CC Pfam: PF01483; P. 1.
 CC Pfam: PF00082; Peptidase_S8; 1.
 CC PRINTS: PR00723; SUBTILASIN.
 CC PROSITE: PS00136; SUBTILASE_ASP; 1.
 CC PROSITE: PS00137; SUBTILASE_HIS; 1.
 CC PROSITE: PS00138; SUBTILASE_SER; 1.
 CC KMW: Hydrolyase; Serine protease; Glycoprotein; Zymogen; Calcium; Signal.
 CC FT SIGNAL 1 27
 CC FT PROPEP 28 110
 CC FT CHAIN 111 753
 CC FT DOMAIN 122 410
 CC FT ACT_SITE 167 751
 CC FT ACT_SITE 167 751
 CC FT ACT_SITE 208 208
 CC FT ACT_SITE 382 382
 CC FT CARBOHYD 401 401
 CC FT CARBOHYD 645 645
 CC FT CONFLICT 23 23
 CC FT SEQUENCE 753 AA; 84146 MW; 95878441BBBD9CD CRC64;
 SO
 Query Match 10.18; Score 71.5; DB 1; Length 753;
 Best Local Similarity 31.28; Ptd. No. 58;
 Matches 20; Conservative 11; Mismatches 26; Indels 7; Gaps 2;
 QY 48 REGIOAGVYIKITDQGVYSHGDHYVYNGKVPYDALFSEELMKRPQOLKADIVNE 107
 DB 153 EKGITGAVYIVDDLENNHTDIYANTPEASTD--FNDNDHDPKRI-----DLTNE 205
 QY 108 VKGG 111
 DB 206 NKHG 209

RESULT 15
 SLAP_ACEKI STANDARD; PRT: 762 AA.
 AC P22258;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE CELL SURFACE PROTEIN PRECURSOR (S-LAYER PROTEIN).
 OS Acetogenium kivi.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Thermotomabacter group; Thermotomabacter.
 OX NCBI_TaxID=2325;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-DSM 2030;
 RX MEDLINE-90036724; PubMed-2681162;
 RA Peters J., Peters M., Lottspeich F., Baumeister W.;
 RT "S-layer protein gene of Acetogenium kivi: cloning and expression in
 RL Escherichia coli and determination of the nucleotide sequence.";
 J. Bacteriol. 171:6307-6315(1989).
 RN [2]
 RP PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
 RC STRAIN-DSM 2030;
 RX MEDLINE-92281680; PubMed-1596358;
 RA Peters J., Rudolf S., Oeschkinet H., Mengele R., Sumper M.,
 RT Kellermann J., Lottspeich F., Baumeister W.;
 RL "Evidence for tyrosine-linked glycosaminoglycan in a bacterial
 RL surface protein.";
 J. Biol. Chem. Hoppe-Seyler 373:171-176(1992).
 RN [3]
 RP DOMAINS.
 RX MEDLINE-94156823; PubMed-8113161;
 RA Lupas A., Engelhardt H., Peters J., Santarius U., Volker S.,
 RA Baumeister W.;
 RT "Domain structure of the Acetogenium kivi surface layer revealed by
 RL electron crystallography and sequence analysis.";
 J. Bacteriol. 176:1224-1233(1994).
 CC -1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
 CC OF PROTEINS WHICH CONT THE SURFACE OF BACTERIA.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
 CC S-LAYER WITH HEXAGONAL SYMMETRY.
 CC -1- PTM: THE CARBOHYDRATE CONTENT OF THIS PROTEIN IS ABOUT 8% WHICH
 CC CORRESPOND TO ABOUT 40 TO 50 SUGAR MOLECULES PER MONOMER. O-LINKED
 CC GLYCANS CONSIST OF GLC, GALNAc AND GLCNAC.
 CC -1- SIMILARITY: CONTAINS 3 S-LAYER HOMOLOGY (SLH) DOMAINS.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC
 CC EMBL: M31069; AAA21930.1; -
 CC PIR: A34355; A34355.
 CC InterPro: IPR001119; -
 CC Pfam: PF00395; SLH; 2.
 CC PROSITE: PS01072; SLH_DOMAIN; 2.
 CC KMW: Signal; Glycoprotein; Repeat; Cell wall; S-layer.
 CC FT SIGNAL 1 26
 CC FT CHAIN 27 762
 CC FT DOMAIN 30 94
 CC FT DOMAIN 95 145
 CC FT DOMAIN 146 204
 CC FT DOMAIN 473 479
 CC FT DOMAIN 625 630
 CC FT CARBOHYD 297 297
 CC FT CARBOHYD 516 516
 CC FT CARBOHYD 520 520
 CC FT CARBOHYD 632 632
 CC
 CC CELL SURFACE PROTEIN.
 CC SLH 1.
 CC SLH 2.
 CC SLH 3.
 CC SER/THR-RICH.
 CC SER/THR-RICH.
 CC O-LINKED (GLC...).
 CC O-LINKED (GLC...).
 CC O-LINKED (GLC...).
 CC O-LINKED (GLC...).

SEQUENCE 762 AA: 82785 MW: 34EC9C784DECA67E CRC64:

Query Match 10.1%; Score 71.5; DB 1; Length 762;

Best Local Similarity 23.8%; Pred. NO. 58; Matches 31; Conservative 19; Mismatches 41; Indels 39; Gaps 7;

QY 21 DNNRVSIVDGSQS-----SOKSENLP-----DOVSQKEGIAEQIV 57
DB 279 DNDVVSFTGSDSVGTVYKNDNNTAKVDNNAVLYNGYLTYSKYVKEGAE----- 333
QY 58 IKITDGGVTSQDHYHYNGKVPY-DALFSEELMKDPNYQKADADIVNEVKGGIIRV 116
DB 334 VTIINNNYLIIVNGS---YDNSTIVNDVQSGDKYLNKDSNTEKGTAVTV-----GAVSKV 386
QY 117 ---DKYIYY 123
DB 387 TDIKANDYIY 396

Search completed: September 26, 2001, 22:34:31
Job time: 235 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 22:32:23; Search time 46.65 seconds

(without alignments)
222.074 Million cell updates/sec

Title: US-09-471-255-2_COPY_15_150

Perfect score: 711
Sequence: 1 IVSLSLCAVALNHRSGENK.....DKRYVYLKDAHADNVRK 136

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: 1: PIR_681:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| No. | Score | Query Match | Length | ID | Description |
|-----|-------|-------------|--------|----|---------------------|
| 1 | 423 | 59.5 | 822 | 2 | hypothetical 92.4K |
| 2 | 85.5 | 12.0 | 721 | 2 | virulence associat |
| 3 | 83 | 11.7 | 814 | 2 | DNA gyrase subunit |
| 4 | 82 | 11.5 | 1946 | 2 | lactococcal (EC 3.4 |
| 5 | 79.5 | 11.2 | 567 | 2 | flagellar m-ring p |
| 6 | 79.5 | 11.2 | 2573 | 2 | hypothetical prote |
| 7 | 78 | 11.0 | 262 | 2 | triacylglycerol 11 |
| 8 | 78 | 11.0 | 438 | 2 | spore cortex-lytic |
| 9 | 78 | 11.0 | 642 | 2 | probable anthranil |
| 10 | 77.5 | 10.9 | 857 | 2 | T04208 |
| 11 | 77 | 10.8 | 1785 | 2 | major merocozite su |
| 12 | 76.5 | 10.8 | 594 | 2 | hypothetical prote |
| 13 | 76 | 10.7 | 875 | 2 | ubiquitin-specific |
| 14 | 75.5 | 10.6 | 312 | 2 | leukocidin chain S |
| 15 | 75.5 | 10.6 | 312 | 2 | leucocidin chain S |
| 16 | 75 | 10.5 | 312 | 2 | hypothetical prote |
| 17 | 75 | 10.5 | 1060 | 2 | vitellinogen conve |
| 18 | 74.5 | 10.5 | 388 | 2 | GPR-binding protei |
| 19 | 74.5 | 10.5 | 836 | 2 | regulatory protein |
| 20 | 74 | 10.4 | 710 | 1 | prolyl oligopeptid |
| 21 | 74 | 10.4 | 710 | 1 | prolyl oligopeptid |
| 22 | 74 | 10.4 | 710 | 1 | prolyl oligopeptid |
| 23 | 73.5 | 10.3 | 1028 | 2 | hypothetical prote |
| 24 | 73 | 10.3 | 113 | 2 | sensorin A - Calif |
| 25 | 73 | 10.3 | 282 | 2 | protein p23 homolo |
| 26 | 73 | 10.3 | 417 | 2 | hypothetical prote |
| 27 | 73 | 10.3 | 710 | 2 | prolyl endopeptida |
| 28 | 72.5 | 10.2 | 1712 | 2 | hypothetical prote |
| 29 | 72 | 10.1 | 205 | 2 | ulcer-associated g |

| | | | | | | |
|----|------|------|------|---|--------|---------------------|
| 30 | 72 | 10.1 | 238 | 2 | A46130 | fiber protein E6 (|
| 31 | 72 | 10.1 | 246 | 2 | S65063 | fiber protein E6 (|
| 32 | 72 | 10.1 | 567 | 2 | G64563 | flagellar basal-bo |
| 33 | 72 | 10.1 | 601 | 2 | T32486 | hypothetical prote |
| 34 | 72 | 10.1 | 1116 | 2 | A35129 | surface layer prot |
| 35 | 72 | 10.1 | 1592 | 2 | A38175 | glucosyltransferas |
| 36 | 71.5 | 10.1 | 316 | 2 | T19475 | hypothetical prote |
| 37 | 71.5 | 10.1 | 753 | 1 | KXMSL1 | proteobact convert |
| 38 | 71.5 | 10.1 | 762 | 2 | A34355 | cell surface prote |
| 39 | 71.5 | 10.1 | 1120 | 2 | S67208 | hypothetical prote |
| 40 | 71.5 | 10.1 | 1714 | 2 | E71609 | Ser/Thr protein kl |
| 41 | 71 | 10.0 | 478 | 2 | A32555 | major merocozite su |
| 42 | 71 | 10.0 | 533 | 2 | A39410 | aldehyde dehydroge |
| 43 | 71 | 10.0 | 3147 | 2 | T18674 | hypothetical prote |
| 44 | 70.5 | 9.9 | 241 | 2 | S65062 | fiber protein E6 (|
| 45 | 70.5 | 9.9 | 318 | 2 | C64445 | conserved hypothe |

ALIGNMENTS

RESULT 1

T46758
hypothetical 92.4K protein - Streptococcus agalactiae

C:Species: Streptococcus agalactiae

C>Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000

C/Accession: T46758

R:Spellerberg, B., Rozdzinski, E., Martin, S., Weber-Heymann, J., Schitzler, N., J

Infect. Immun. 67, 871-878, 1999

A:Title: Lmb, a protein with similarities to the Lrai adhesin family, mediates attach

A:Reference number: 224091; MUID:99115568

A:Accession: T46758

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-822 <SEP>

A:Cross-references: EMBL:AF062533; NID:94249622; PIDN:RAD31379.1; PID:94249624

A:Experimental source: strain R268

C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 59.5%; Score 423; DB 2; Length 822;
Best Local Similarity 55.1%; Pred. No. 7.3e-30;
Matches 76; Conservative 26; Mismatches 34; Indels 2; Gaps 1;

OY 1 IVSLSLCAVALNHRSGENKNNRVSVDGSSQSKS--ENLTPDVQSKGIGQAEQIV 58
DB 15 ILAHHIGSYQKHHMGATKDNQATYDSSKGVKAKTKTKMDQISAEGISAEQIV 74
OY 59 KITDGYTSHGDHYHYNGRVPYDALFSEELMKDPYQKADIVNEVKGIIKVDG 118
DB 75 KITDGYTSHGDHYHYNGRVPYDALFSEELMKDPYQKADIVNEVKGIIKVDG 134
OY 119 KYVYVYLKDAHADNVRK 136
DB 135 NYVYVYLKPKSKRKNIPTK 152

RESULT 2

C82939
virulence associated protein, exoribonucleases U0057 [Imported] - Ureaplasma urealyticum

C:Species: Ureaplasma urealyticum

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C/Accession: C82939

R:Glass, J.I., Leikowitz, E.J., Glass, J.S., Helner, C.R., Chen, E.Y., Cassell, G.H.

submitted to GenBank, February 2000

A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a

A:Reference number: A82870

A:Accession: C82939

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-721 <GLA>

A:Cross-references: GB:AE002105; GB:AE222894; NID:96899003; PIDN:AAF30462.1; GSPDB:G

A:Experimental source: serovar 3; biovar 1

Db 235 ALBOLKRNENLEKINVLAPIVGKKNKRYNAEDFSQKSTKETEPNNVVR 294
 QY 46 -----SOKESIOAEQIVITITOGYTSNGDHYHNGVPPIDALFSEELMKDPYOL 99
 Db 295 SEQLLEKKEGAPKRYG---GVPGVYVSNIGPVGGLDNKEPE---KYEKSONTYNEYV 347
 QY 100 KDAIVNEVKG---IIVDKRYVYLKDAHA 130
 Db 348 -GRTISEIRGEFTLVRLNAAVVDGKRYKIALEDGANA 384

RESULT 6

14
 Chemical protein PRB0460c - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
 C:Accession: D71614
 R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 R:Perren, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
 Science 282, 1126-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; M01D:99021743
 A:Accession: D71614
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2573 <GAR>
 A:Cross-references: GB:AE001366; GB:AE001362; NID:93845188; PIDN:AAC71881.1; PID:9384518
 A:Experimental source: clone 3D7
 C:Genetics:
 A:Gene: PRB0460c

Query Match 11.28; Score 79.5; DB 2; Length 2573;
 Best Local Similarity 20.88; Pred. No. 1.2e+02;
 Matches 31; Conservative 25; Mismatches 38; Indels 55; Gaps 6;
 QY 12 NOHRSQENKNNRVSIVDGSOSQKSE-----NLPPD-----QV 45
 Db 2421 NDNNDNNNDNNNNSHLAFQNRTOGETTPTNNINNTDICEKGNKYTSNVNNINNM 2480
 QY 46 SOKESIOAEQIVITITOGYTSNGDHYH-----YNGKVPYDA 84
 Db 2481 TCKSEVEVEEIIQTKNRKF---HNIELEKHCYDLEFKRKLNTYRNTYKKNKIINC 2537
 QY 85 LFESEELMKDPYOLKDAIVNEVKGYY 113
 Db 2538 LITNKNT-----FOYKENDIVNKYKQIF 2561

RESULT 7

JC4109
 triacylglycerol lipase (EC 3.1.1.3) 1 - Mycoplasma mycoides subsp. mycoides
 N:Alternate names: lipase
 C:Species: Mycoplasma mycoides subsp. mycoides
 C>Date: 23-Jul-1995 #sequence_revision 13-Oct-1995 #text_change 07-Dec-1999
 C:Accession: JC4109
 R:Rawadi, G.; Lalanne, J.L.; Roulland-Dussoix, D.
 Gene 158, 107-111, 1995
 A:Title: Cloning and characterization of the lipase operon from Mycoplasma mycoides subsp.
 A:Reference number: JC4109; M01D:95309706
 A:Accession: JC4109
 A:Molecule type: DNA
 A:Residues: 1-262 <RAM>
 A:Cross-references: GB:U17036
 A:Note: The authors translated the codon TGA for residue 162 and 236 as Trp
 C:Comment: This enzyme, a serine esterase, is widely distributed throughout animals, pla
 ty acids, and hydrolyses ester bonds of triacylglycerols to yield free fatty acids, di
 C:Genetics:
 A:Genetic code: SGC3
 C:Superfamily: triacylglycerol lipase 1
 C:Keywords: carboxylic ester hydrolase
 F:90-94/Region: conserved site

Query Match 11.0%; Score 78; DB 2; Length 262;
 Best Local Similarity 26.4%; Pred. No. 9.9;
 Matches 29; Conservative 15; Mismatches 50; Indels 16; Gaps 4;

QY 9 YALNHRSQENKNNRVSIVDGSOSQKSENLPPDVSQREGIOAEQIV--KITDGYV 66
 Db 5 YDNYVFKNNNNNNENIIFVHGYNSS-----PRFEYLKNIQDDIIMHYNFQDDIV 57
 QY 67 TSHGDHYHNGVPPIDALFSEELMKDPYOLKDAIVNEVKGYYIIV 116
 Db 58 KPYKDH-----KTVWEG-FAOLLHIEFQNIKNVAIGHSMGGVYIS 100

RESULT 8

140824
 spore cortex-lytic enzyme precursor - Clostridium perfringens
 C:Species: Clostridium perfringens
 C>Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 15-Oct-1999
 C:Accession: I40824; PC2363
 R:Myata, S.; Moriyama, R.; Miyahara, N.; Makino, S.
 Microbiology 141, 2643-2650, 1995
 A:Title: A gene (slec) encoding a spore-cortex-lytic enzyme from Clostridium perfring
 A:Reference number: I40823; M01D:96036223
 A:Accession: I40824
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-438 <RES>
 A:Cross-references: GB:D45024; NID:9940391; PIDN:BA08081.1; PID:9940393
 R:Myata, S.; Moriyama, R.; Sugimoto, K.; Makino, S.
 Biotech. Biotechnol. Biochem. 59, 514-515, 1995
 A:Title: Purification and partial characterization of a spore cortex-lytic enzyme of
 A:Reference number: PC2363; M01D:95252603
 A:Accession: PC2363
 A:Molecule type: protein
 A:Residues: 150-164 <MIT>
 C:Genetics:
 A:Gene: slec

Query Match 11.0%; Score 78; DB 2; Length 438;
 Best Local Similarity 28.6%; Pred. No. 18;
 Matches 26; Conservative 10; Mismatches 27; Indels 28; Gaps 4;

QY 58 KITDGYVTSNGDHYHNGVPPIDALFSEELMKD-----PNYQLDA- 102
 Db 278 INVCCPMNTQWGSRYLGDDEKGVYDILTSFYGDLEKSAKRYGSPRSYPTLTATGY 337
 QY 103 -----DIVNEVKG---IIVDKRY 120
 Db 338 SGEPVRYIOELNISRATPLIPKIAVDGKY 368

RESULT 9

DB1401
 Probable flagellar hook-associated protein Cj0548 [imported] - Campylobacter jejuni (C)
 C:Species: Campylobacter jejuni
 C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
 C:Accession: DB1401
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chli
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Be
 Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
 A:Reference number: AB1250; M01D:20150912
 A:Accession: DB1401
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-642 <PAR>
 A:Cross-references: GB:AL139075; GB:AL111168; NID:96967817; PIDN:CA075184.1; PID:9696
 A:Experimental source: serotype O2, strain NCTC 11188
 C:Genetics:
 A:Gene: fldD; Cj0548

Query Match 11.0%; Score 78; DB 2; Length 642;
 Best Local Similarity 28.3%; Pred. No. 29;
 Matches 28; Conservative 16; Mismatches 39; Indels 16; Gaps 4;

QY 21 DNNRVSVDGSSQSKSENLPDVSQKQIAEQIVIKITDQGYVSH---GDHYHY 76
 DB 84 DNPASLVNNGVALQSNINVTQAGKDYQSKGLA---NDGFRVNAQLNGTDLTFES 140
 QY 77 NGKVPYDALFSEELMKDPNTQKLD-ADIVNEVKGYYII 114
 DB 141 NGK-----EYTVYDKNTTYTDLADKINEASGGEIV 171

RESULT 10

T04208
 Probable anthranilate phosphoribosyltransferase (EC 2.4.2.18) TSC23.40 - Arabidopsis th
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 11-Jun-1999
 C:Accession: T04208
 R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
 submitted to the Protein Sequence Database, March 1999
 A:Reference number: Z15361
 A:Accession: T04208
 A:Molecule type: DNA
 A:Status: 1-857 <BEV>
 A:Cross-references: EMBL:AL049500;
 A:Experimental source: cultivar Columbia; BAC clone TSC23
 C:Genetics:
 A:Map position: 4
 A:Introns: 150/3; 280/3; 391/2
 A:Note: TSC23.40
 C:Keywords: glycosyltransferase; pentosyltransferase

Query Match 10.9%; Score 77.5; DB 2; Length 857;
 Best Local Similarity 21.8%; Pred. No. 47;
 Matches 32; Conservative 25; Mismatches 39; Indels 51; Gaps 7;

QY 10 ALN-QHRSQEKDNNRVSVDGSSQSKSENLPDVSQKQIAEQIVIKITDQGYVTS 68
 DB 3 AMNVEHRS---DKRHVY-MLPNSAQEHQHPQGPNOSSSLAAEQ-----DN 46
 QY 69 HGDRHYHYNGKVPYDALFSE-----ELMKDPNTQKLDADIVNEVK 109
 DB 47 HNEHHHHVPRKQYDEKSEPARPSKLYVHAHSTASAPADALKEYSPLHGGGRV---- 102
 QY 110 GGYIIR-----VDGRYYVYLK 125
 DB 103 GGRVYHDKDTATSTYDVERMYFLYVR 129

RESULT 11

T50182
 merizolite surface antigen precursor - Plasmodium chabaudi chabaudi
 C:Species: Plasmodium chabaudi chabaudi
 C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-Aug-1999
 C:Accession: A45546
 R:Deleersnijder, W.; Hendrix, D.; Bendalman, N.; Hanegreets, J.; Briljs, L.; Hamers-Caster
 Mol. Biochem. Parasitol. 43, 231-244, 1990
 A:Title: Molecular cloning and sequence analysis of the gene encoding the major merizolite
 A:Reference number: A45546; MID:91218805
 A:Accession: A45546
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1785
 A:Cross-references: GB:M34947; NID:g160597; PID:g160598
 C:Superfamily: major merizolite surface antigen
 C:Keywords: surface antigen

Query Match 10.8%; Score 77; DB 2; Length 1785;

Best Local Similarity 21.1%; Pred. No. 13e+02;
 Matches 23; Conservative 28; Mismatches 36; Indels 22; Gaps 5;

QY 21 DNNRVSVDGSSQSKSEN-LTPQVSQKQIAEQIVIKITDQGYVTSQDHH--YVN 77
 DB 1104 DGNNTQIDPFRSKKEILNLTPEKVNQ-----LYLDIAHKEISEHYNNRYK 1152
 QY 78 GRVYVDALFSEELMKDPNTQKLDADIVNEVKGYYIK---VDGRYYV 122
 DB 1153 YKLEIRLYCHQHEQIEAANKVKEISYK-----SRLAKRKYYNGTYY 1197

RESULT 12

B64018
 hypothetical protein HI1004 - Haemophilus influenzae (strain Rd KW20)
 C:Species: Haemophilus influenzae
 C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 29-Sep-1999
 C:Accession: B64018
 R:Gieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
 ; Ocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Weidman
 ; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
 Science 269, 496-512, 1995
 A:Authors: Guehm, C.L.; McDonald, L.A.; Small, K.V.; Frazer, C.M.; Smith, H.O.; Vente
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; MID:95350630
 A:Accession: B64018
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-594 <TIGR>
 A:Cross-references: GB:U32781; GB:I42023; NID:g1574028; PIDN:AMC22665.1; PID:g1574034
 C:Genetics:
 A:Start codon: GTG
 C:Superfamily: hypothetical protein HI1004

Query Match 10.8%; Score 76.5; DB 2; Length 594;
 Best Local Similarity 21.0%; Pred. No. 36;
 Matches 37; Conservative 29; Mismatches 67; Indels 43; Gaps 6;

QY 1 IVSISLCAVYALNDRSQEN---KDNRRVSVDGSSQSKSENLPDVSQKQIAEQI 56
 DB 168 LATLSLADENAKQSVSDDEIKTYEANKQSPVOPEQVQYIDLSADNISRNLOVDVEI 227
 QY 57 VIRITDQ--GYVSHGDRHYHYNGKVPYDALFSEELMKDPNT---QLDADIVNEVK 110
 DB 228 AQTQYDKAKQPMQHLAHIOFANEQ---DAKVAEEELQKAMPADVAKASLDKISENG 284
 QY 111 GYT-----IKYDGKYYVL---KDAHADNVYRK 136
 DB 285 GDLGAVNENELPRAFEDEAAALQVGQSPINVDGNVHYLVQERKAQSLNENKAQ 340

RESULT 13

T50182
 ubiquitin-specific proteinase homolog [imported] - fission yeast (Schizosaccharomyces
 C:Species: Schizosaccharomyces pombe
 C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
 C:Accession: T50182
 R:Radcock, K.; Churcher, C.M.; Wood, V.; Bartrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, February 2000
 A:Reference number: Z25044
 A:Accession: T50182
 A:Status: preliminary; translated from GB/EMBL/DDBB
 A:Molecule type: DNA
 A:Residues: 1-875 <BAD>
 A:Cross-references: EMBL:AL138854; PIDN:CA872233.1; GSPDB:GN00066; SPDB:SPAC2363.08C
 A:Experimental source: EMBL: strain 97h(-); cosmid c2363
 C:Genetics:
 A:Gene: SPDB:SPAC2363.08C
 A:Map position: 1

Query Match 10.7%; Score 76; DB 2; Length 875;

Best Local Similarity 21.0%; Pred. No. 65;
Matches 30; Conservative 27; Mismatches 46; Indels 40; Gaps 5;
QY 2 VSLSLAYALNORSEKNDNR-----VSYVDG-----SSQSKSE-----38
DB 356 VNTDLSLSPNSHTSDNEDNEDYVSLSVSEETEDITYSKLSQSDSPSOHQHDFL 415
QY 39 --MLTPDQVQKFGIOAEQIVIKITDQGYTSHG-----DHYHYNGKVPY 82
DB 416 PANSFPLAASSTSLPSSSEILDSSSDKGOQVFGQHEVACTNSFEDPNSHFNVSNSNHE 475
QY 83 DALFSEELMKDPNYQKADIV 105
476 EASPKKEVL-KSPQFQRRSLDL 497

RESULT 14
T00160
leukocidin chain S precursor - staphylococcus aureus phage phi PVL
N/Alternate names: protein 27
C/Species: Staphylococcus aureus phage phi PVL
C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
C/Accession: T00160
R/Kaneho, J.; Kilmura, T.; Kawakami, Y.; Tomita, T.; Kamio, Y.
Biosci. Biotechnol. Biochem. 61, 1960-1962, 1997
A/Title: Panton-Valentine leukocidin genes in a phage-like particle isolated from mitomy
A/Reference number: Z14119; M01D:98067870
A/Accession: T00160
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-312 <KAN>
A/Cross-references: EMBL:AB009866; PIDN:BA31900.1
C/Genetics:
A/Genes: luks
A/Function:
A/Description: luks and luks cooperatively and strongly lyse rabbit erythrocytes besides
C/Superfamily: leukocidin
C/Keywords: hemolysis; toxin

Query Match 10.6%; Score 75.5; DB 2; Length 312;
Best Local Similarity 23.0%; Pred. No. 20;
Matches 26; Conservative 27; Mismatches 33; Indels 27; Gaps 6;
QY 10 ALNQHRSQENKDNRRVSYDGSQSSQSKSENLPDQVSQKEGIG-----AEQIVIK 59
DB 20 ATSFHESK--ADNNINENIGDAEVYKRTEDTSSDKNGVYQNIQFDFVKDKYKNDALILK 77
QY 60 ITDQGYTSHGDHYHYNG-----KVPYDALFSEELMKDPNYQKADIVN 106
DB 78 M--QGFINSKTYTYNKNPDHIRMKRPQ--YNIGLKTNDPN-----VDLIN 121

RESULT 15
S32211
leucocidin chain S - Staphylococcus aureus
C/Species: Staphylococcus aureus
C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 17-Mar-2000
C/Accession: S32211
R/Prevost, G.; Supersac, G.; Piemont, Y.
submitted to the EMBL Data Library, March 1993
A/Description: Staphylococcus aureus encodes two types of symergohymenotropic toxins: P
A/Reference number: S32211
A/Accession: S32211
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-312 <PRE>
A/Cross-references: EMBL:X72700
C/Genetics:
A/Genes: luks
C/Superfamily: leukocidin
C/Keywords: toxin

Query Match 10.6%; Score 75.5; DB 2; Length 312;
Best Local Similarity 23.0%; Pred. No. 20;
Matches 26; Conservative 27; Mismatches 33; Indels 27; Gaps 6;
QY 10 ALNQHRSQENKDNRRVSYDGSQSSQSKSENLPDQVSQKEGIG-----AEQIVIK 59
DB 20 ATSFHESK--ADNNINENIGDAEVYKRTEDTSSDKNGVYQNIQFDFVKDKYKNDALILK 77
QY 60 ITDQGYTSHGDHYHYNG-----KVPYDALFSEELMKDPNYQKADIVN 106
DB 78 M--QGFINSKTYTYNKNPDHIRMKRPQ--YNIGLKTNDPN-----VDLIN 121

Search completed: September 26, 2001, 22:32:25
Job time: 309 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 22:30:36, Search time 32.85 Seconds

(without alignments)
1083.454 Million cell updates/sec

Title: US-09-471-255-2

Perfect score: 5406
Sequence: 1 MRFKRYIAAGSAVTVSL.....IELRLPSEGVTKKNSDFIA 1039

oring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 227.5 | 4.2 | 1701 | 1 | MSPI_PLAFM |
| 2 | 226.5 | 4.2 | 1701 | 1 | MSPI_PLAFM |
| 3 | 205.5 | 3.8 | 1183 | 1 | CNA_STAU |
| 4 | 205.5 | 3.8 | 1726 | 1 | MSPI_PLAFM |
| 5 | 205.5 | 3.8 | 1726 | 1 | MSPI_PLAFM |
| 6 | 203.5 | 3.8 | 1637 | 1 | MRSP_STAU |
| 7 | 203.5 | 3.8 | 1849 | 1 | IGA4_HAEN |
| 8 | 189 | 3.5 | 1682 | 1 | MSPI_PLAFM |
| 9 | 186.5 | 3.4 | 1570 | 1 | P3K1_DICD1 |
| 10 | 185 | 3.4 | 2194 | 1 | SC16_YEAST |
| 11 | 184.5 | 3.4 | 1466 | 1 | SPA2_YEAST |
| 12 | 179.5 | 3.3 | 1636 | 1 | BUD3_YEAST |
| 13 | 177 | 3.3 | 1702 | 1 | IGA2_HAEN |
| 14 | 176.5 | 3.3 | 1532 | 1 | HMW1_MYCGE |
| 15 | 176.5 | 3.3 | 1532 | 1 | IGA_NEIGO |
| 16 | 176.5 | 3.3 | 1658 | 1 | YMG7_YEAST |
| 17 | 175.5 | 3.2 | 2578 | 1 | NOM1_YEAST |
| 18 | 174.5 | 3.2 | 1577 | 1 | HLA_PROMI |
| 19 | 174.5 | 3.2 | 1807 | 1 | VTAG_XENIA |
| 20 | 174.5 | 3.2 | 2485 | 1 | PTND_HUMAN |
| 21 | 173 | 3.2 | 1694 | 1 | IGA0_HAEN |
| 22 | 173 | 3.2 | 1781 | 1 | AKAC_HUMAN |
| 23 | 173 | 3.2 | 1790 | 1 | USO1_YEAST |
| 24 | 172.5 | 3.2 | 1142 | 1 | GIN4_YEAST |
| 25 | 172 | 3.2 | 817 | 1 | YGA4_YEAST |
| 26 | 172 | 3.2 | 2869 | 1 | RBP1_PLAFM |
| 27 | 171 | 3.2 | 710 | 1 | LT78_ARATH |
| 28 | 169.5 | 3.1 | 1167 | 1 | SCPA_STRPY |
| 29 | 169 | 3.1 | 1744 | 1 | TANA_XENIA |
| 30 | 168.5 | 3.1 | 818 | 1 | PKR2_YEAST |
| 31 | 168.5 | 3.1 | 1435 | 1 | EBAL_PLAFM |
| 32 | 168.5 | 3.1 | 1435 | 1 | LTBL_YEAST |
| 33 | 168 | 3.1 | 1630 | 1 | MSPI_PLAFM |

| | | | | | | |
|----|-------|-----|------|---|------------|--------------------|
| 34 | 168 | 3.1 | 3924 | 1 | ANK2_HUMAN | 001484 homo sapien |
| 35 | 167 | 3.1 | 1639 | 1 | MSPI_PLAFM | P04933 plasmodium |
| 36 | 165.5 | 3.0 | 1256 | 1 | ATL_STAU | P52081 staphylococ |
| 37 | 164.5 | 3.0 | 1164 | 1 | BAG_STRAG | P27951 streptococ |
| 38 | 164.5 | 3.0 | 1189 | 1 | YTH6_YEAST | P47035 saccharomyc |
| 39 | 164 | 3.0 | 1222 | 1 | S160_YEAST | P06105 saccharomyc |
| 40 | 164 | 3.0 | 1679 | 1 | YMG9_YEAST | P21979 streptococ |
| 41 | 163.5 | 3.0 | 1528 | 1 | SPA3_YEAST | P43385 haemophilus |
| 42 | 163.5 | 3.0 | 1545 | 1 | IGA3_HAEN | 009904 schizosacch |
| 43 | 163 | 3.0 | 1159 | 1 | N124_SCHPO | P40477 saccharomyc |
| 44 | 163 | 3.0 | 1460 | 1 | N159_YEAST | P16053 gallus gall |
| 45 | 162.5 | 3.0 | 857 | 1 | NFM_CHICK | |

ALIGNMENTS

| RESULT | ID | MSPI_PLAFM | STANDARD | PRT | 1701 AA. |
|--------|--|------------|----------|-----------------------------------|----------|
| AC | P08569 | MSPI_PLAFM | | | |
| DT | 01-AUG-1988 (Rel. 08, Created) | | | | |
| DT | 30-MAY-2000 (Rel. 39, Last sequence update) | | | | |
| DT | 30-MAY-2000 (Rel. 39, Last annotation update) | | | | |
| DE | MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS) | | | | |
| DE | (PMMSA) (P190). | | | | |
| GN | MSPI-1 | | | | |
| OS | Plasmodium falciparum (isolate mad20 / Papua New Guinea). | | | | |
| OC | Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. | | | | |
| ON | NCBI_TaxID=70153; | | | | |
| OX | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RX | MEDLINE=88011243; Pubmed=3079521; | | | | |
| RA | Tanabe K., Mackay M., Goman M., Scalfe J.G. | | | | |
| RT | "Allelic dimorphism in a surface antigen gene of the malaria parasite | | | | |
| RT | Plasmodium falciparum." | | | | |
| RL | J. Mol. Biol. 195:273-287(1987). | | | | |
| RN | [2] | | | | |
| RP | REVISIONS TO 1403; 1569 AND 1629. | | | | |
| RA | Tanabe K.; (JAN-1999) to the EMBL/GenBank/DBJ databases. | | | | |
| RL | Submitted | | | | |
| RN | [3] | | | | |
| RP | SEQUENCE OF 1-115 FROM N.A. | | | | |
| RX | MEDLINE=86136024; Pubmed=3004972; | | | | |
| RA | Mackay M., Goman M., Bone N., Hyde J.E., Scalfe J., Certa U., | | | | |
| RT | Stuenkelberg H., Bujard H. | | | | |
| RT | "Polymorphism of the precursor for the major surface antigens of | | | | |
| RT | Plasmodium falciparum merozoites: studies at the genetic level." | | | | |
| RL | EMBO J. 4:3823-3829(1985). | | | | |
| CC | -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR | | | | |
| CC | (POTENTIAL) | | | | |
| CC | -1- PFM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42 | | | | |
| CC | KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF | | | | |
| CC | MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT. | | | | |
| CC | THIS SWISS-PROT entry is copyright. It is produced through a collaboration | | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | | |
| CC | or send an email to license@isb-sib.ch). | | | | |
| CC | ----- | | | | |
| CC | EMBL: X05624; CAA29112.1; | | | | |
| DR | PIR: A26868; A26868. | | | | |
| DR | PIR: B25120; B25120. | | | | |
| DR | Interpro: IPR000561; | | | | |
| DR | Pfam: PF00008; EGF_1. | | | | |
| KW | Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein; | | | | |
| KW | Transmembrane; GPI-anchor. | | | | |
| FT | SIGNAL | 20 | 1701 | POTENTIAL. | |
| FT | CHAIN | 1 | 19 | MEROZOITE SURFACE PROTEIN 1. | |
| FT | CARBOHYD | 110 | 110 | N-LINKED (GLCNAC. .) (POTENTIAL). | |

Query Match 4.28; Score 227.5; DB 1; Length 1701;
 Best Local Similarity 20.28; Pred. No. 0.005;
 Matches 250; Conservative 176; Mismatches 405; Indels 405; Gaps 66;

Query 1 MFSSKTYAGSAVYSLCAVALNORSOEN---KDNRVSYVDSGSGSSENLTP 56
 500 MKNFNNF---DKDYVDKFFSARYYVNEKRYNKFSSNNSTYV---QKTKALSTLE 553
 57 DVSQKESIQAEQYVITDGGYVTSQGDHYHYNGRPVDA---LFSEEL----- 104
 554 D-YSLRKGI-----SERDFHYHYTLKTLGLADIKLLEIKSEENKLE 596
 105 ---LMDPYOLKADIVN-EYKGGYII-KVDS--KYVYVLDAAHADVVR----- 149
 597 KNFGGLHSANASLEVSIVYALVOVKVLLIKRIEDLRIELFLKNAOLKDSIHVNITKP 656
 150 ---KDEINRKO-----EHVKNKENVSNVA---VARSGRHTTNDG--- 185
 657 QNKPEPYLYLKKKEVDKKEIFKRVMDLKEQAVLSITQPLVASE---TTEDGHS 713
 186 ---YVNPADIIEDTGAAYVPRGCHYHYI---PRSDLSASELAARAKHIA 230
 714 THTSOSGETEVEETEETEET-----VGHHTTVTITLPRRESAPKRVKENSIE 765
 231 GKNMOPSLSYSSPAASNNQSVAKSGSTRANK-----SENLSLKL 275
 766 HK-----SNDNSQALKTYVLYLKLDELFLKSYICHKYLIVSNSSMOQKLEV 812
 276 YD-SPSAQRYSESDGLVFPKIIISRTPNGVALPHG-----DHYHFIYSK----- 320
 813 YNLPRPEEKELKS---CDPLDLFNQNNIPANYSLYDSNNDOLOLFLFELYQKEMITY 866
 321 LSALEER-----IARVNPISGTSVST-NAPRNEVYSSLSISNPSLTTSKELSA 373
 869 LHKLEENHKLLEBOKOITGTSSTSPGNTVYTAQSAHNSQSQNSAS---TNT 925
 374 SDGYTFNPNDIVETATAYIRHGDHFIYIPKSNQI-GOPTLPR--NSLATPSPSLP 430
 926 QNG-----VAVSSGPAVVEESHPDPLTVSISMDLKGIVSLMKGKTVAPN-LLTST 977
 431 GTSHEKEDEGDFGDANRIIAEDESDFVMSHGHNHYFFKRD-----LLEEOIKAAQKH 494
 978 -TEMEKTYEN-----LKNNDTYF-----NDIKQFVANSNSVITGLFETQKNLANDE 1024
 485 LEVYKTS-----HN-----GLDSLSHEDDYQENNAEKMD--LDKRIEKIAGIKROY 530
 1025 IKTKDLQSLDFDLYNNYKTLKDLRFKFKKELSGDKMKIKLTLKQLSKLSLNLNPH 1084
 531 GYKAE-SIVVNEKKNATII-----YPHGDHHAAPIDIEHNPVGI 567
 1085 NYLQNEFVFNKKKEAETIAETENTLENTLILKHYKGLVYTYNGE---SSPLTTLSEVSI 1141
 568 --GHSNRYELFRPEEGVAKKEGKRYTYGE-----LTNVNVLKKNSTFNNOFTLA 617
 1142 QTEBNYANLEKFRALSKIDKLDNLDLHGKKLISFLSSGLHHLITELK-EVINKNVT-- 1198
 618 NGQKRVSFSPPELEKTLGIMLVKL---ITPOKYLEKYSGVYFGEGVGNINAFELDOP 674

DB 1199 -----GNSPSENNKR--VNEALKSENEFPPEAKVTVVTP-----POP 1234
 QY 675 YLPOTFEYRTI-----ASKDYPEVSYDQTFVPTSLAYKMASQTIFFEPHADGTYLRNP 729
 DB 1235 DVTSPSLSVKRYSGSGSKETQIPTSQ-----SLTELOOVOLONTDEEDSLVYLP 1288
 QY 730 QFAVPRGTD-----ALVRYDFEFGNAVYENNYKVEIKLPKLNQGTTPRAGN 779
 DB 1289 IFGSESDDEVLDQVVTGEALSVYMDNITSG--FENEEDVLYLK-PL---AGVYRLSKK 1341
 QY 780 KIP---VTFMANA-----YLD-----NOSTYVYEPILKENQTD 811
 DB 1342 QIERNITTFNINLMDIILNSRIKKRRYFLVLESDLMQFKHISSENYIIE-DSFILNSEQ 1400
 QY 812 KPSLLPQFKNRK-----AOENSKLDEYV-----EERTSEKVEKESLSEGNS 854
 DB 1401 KNTLISKRYKTESVENDIKFAQGISYEVKALYDDESIRKVIKEREKPPSSPT 1460
 QY 855 TSNSTLEEVPTVDVQEKVAKFA-----ESYCKLENYLFNM----- 891
 DB 1461 TPFS-----PAKDEQAKKESFLPFLNIEFLYNNVKNKIDDIYLNKAKINDCNVERKE 1515
 QY 892 -----DGTIELYLPSEVYIKRMADETEAPQNGENRPSNGK-VSTGTV 936
 DB 1516 AHVYTKLSDLKALDKIDIF-----KNTNDEFAIKKLINDTRKMDLGLSTGLV 1567
 QY 937 ENOPT-----ENKRPDSLEAPNE--KRYKENS 963
 DB 1568 QNFPNTIISKLIEGKFQDMNTISOHCVKRQCPENS 1603

RESULT 2

MSPL PLAF 2
 ID MSPL PLAF STANDARD; PRT; 1701 AA.
 AC P13819;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MERZOITE SURFACE PROTEIN 1 PRECURSOR (MERZOITE SURFACE ANTIGENS) (PMSA).
 GN MSP-1.
 OS Plasmodium falciparum (Isolate FC27 / Papua New Guinea).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=88142999; PubMed=2449612;
 RX Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G., Brown G.V., Anders R.F., Kemp D.J.;
 RA "Variation in the precursor to the major merozoite surface antigens of Plasmodium falciparum."
 RT Mol. Biochem. Parasitol. 27:291-302(1988).
 RL Mol. Biochem. Parasitol. 27:291-302(1988).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (POTENTIAL).
 CC -1- PTM: MERZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42 KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MERZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).
 CC CC
 CC EMBL: M19143; AAA29653.1;
 DR PIR: A54498; A54498.
 DR InterPro: IPR000561;
 DR Pfam: PR00008; EGF_1.
 KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.

FT CHAIN 20 1701 MEROZOTTE SURFACE PROTEIN 1.
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 607 607 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SEQUENCE 1701 AA; 193719 MW; 3920875E73D38552 CIRC64;

Query Match 4.28; Score 226.5; DB 1; Length 1701;
 Best Local Similarity 20.28; Pred. No. 0.0055;
 Matches 250; Conservative 175; Mismatches 406; Indels 405; Gaps 66;

QY 1 MKFSKTIAGSANTVTSLSICAYALNHRSGEN---KDNRRSYVDGSSOSKSENITP 56
 DB 500 MKFNNNF---DKDYVDKIFSAKYTYNEKORYNKFSSNNSYVNV---QRIKALSTLE 553
 QY 57 DVSQKREGIAEQIVITKTDGYVTSQGDHYHYNGKVPYDA---LRFSEEL 104
 DB 554 D-YSLRKGIT---SEKDNHYITLTGTGLADIKLITEKSKSENKILE 596
 QY 105 ---LAKDPYOLKADIVN-EYKGGYII-KVDG---KYVYLLDAAHADNVR----- 149
 DB 597 KNFGGLHSAASLEVSQIVLQYKVLKIKIDRLKIELFLNAQKDSIHVPIYKP 656
 QY 150 ---KDEINRKO-----EHVKNERNVSNVA---VANSQGYTTNDG--- 185
 DB 657 QNKPEPYLLIKKEVVKLKEFIPKVKDKLKEQAVLSITQPVASE---TLEDGHS 713
 QY 186 ---YVNPADIEDTGNAYIYPHGHYHI-----PKSDLSASELAAKAHIA 230
 DB 714 THYSQSETEVETEETEET-----VGHYTYITLPPREBSAPKVKYVENSIE 765
 QY 231 GKNMPOSLSYSTASDNTQSVAKGSTSKPANK-----SENLSLKL 275
 DB 766 HK-----SNDNSQALTKTYTLKLDLFLKSTYCHYTYLVNSNMQKILEV 812
 QY 276 YD-SPSAQRYSQGLVDFPAKIISRPNGVAPHG-----DHYHETPYSK--- 320
 DB 813 YNLPRPEENELKS---CDPLDLFNIONNIPAMYSITDSNNIDLOHLEFELYOKEMITY 868
 QY 321 LSALEEK-----IARWPIGSGTGYST-NAKPNEVVSISGISNPSLITRELSA 373
 DB 869 LHKLEENHKKLEBOKQITGSTSTSPGTYVTAAQSAVHSNQNQSNASS---TNT 925
 QY 374 SDGYIFNKDIYEETATAYIYRHGDHFHYIPKSNQI-GQPLFPN---NSLAPSPSLPINP 430
 DB 926 QNG-----VAVSSGPAVVESSHDPPLVLSINDLQIVSLNMGNTKTPNP-LTIST 977
 QY 431 GTSHEKHEDGYGDANFIIADESGFYVSHGDNHFFKFD-----LREQIKAAOKH 484
 DB 978 -TEMKEFEEN-----ILKNNDYF---NDDIKQFVKSNSKYITGLTEFOKALNDE 1024
 QY 485 LEEVTS-----HN-----GLDLSHEDQYPGNAKEMKD---LDKRIEKIAGIMQY 530
 DB 1025 IRLKLDYQLQSLDLYNKYKLLKRLFNKKELGQDKMOKIKLTLKEQLESKLSLNPH 1084
 QY 531 GVKRE-STVNNKRNKAIIT-----YPHGDHHADEIDHEKRPYGI 567
 DB 1085 NTLQNFSTFENKKAELAELENLENLKILKHKGKLVKTYNGS---SSPLKLTSEVSI 1141
 QY 568 ---GSHSNVTELEKPEGVAKREKGNVYNGE-----LTVNVNLLKNSFTFNQNTFLA 617
 DB 1142 QREDVYANLEKFRALSKIDKLNQNLHGKKKLSFSLSSGLHLLTELEK-EVTKNNNTY-- 1198

QY 618 NGQKRVSPFPELEKRLGIMLVKL---ITPDGKVLKSYGNVPGVGNINAFELDOP 674
 DB 1199 -----GNPSSENNK--VNEALKSTENFLPEKATVTVTP-----POP 1234
 QY 675 YLPQGTFRYIT-----ASKDYPEVSYDGTFTVPISLAVKASQTIFFPHAGDYLRVNP 729
 DB 1235 DVPSPLSVRVSGSGSTKEETQIPIISG-----SLTELOQVQVQANDEEDSLVYLP 1288
 QY 730 QFVAPKQTD-----ALVRVDFEFGNAVLENYKVEIKLPIPKINQGTTRAGN 779
 DB 1289 IFGESENDGYLDQVYVGEALSVTMDNLSG--FENEVDYIYK-PL-----AGYRSILK 1341
 QY 780 KIP---VTNANA-----YLD-----NOSTIYEPFAEKENQTD 811
 DB 1342 QIKENITTFMLNLDIINSRLKRRKRYFLVDESIDMOFKHSNKEYITE-DSFLLISQ 1400
 QY 812 KPSLTPQFKNK-----AOENSLDERV-----EPPTSKVEKEKSTGNS 854
 DB 1401 KNTLLKSKYKIKESVENDIKFADQGISYIEKVLAKYDOLLSIKYIKEREKPPSPPT 1460
 QY 855 TSNSTLEVPVDPQEKVAKFA-----ESYQKLENVLFNM----- 891
 DB 1461 TPFS-----PAKDQKKESTFLPFLNIETLYNNLVNKTIDVILNKKAKINDGVENKE 1515
 QY 892 -----DGTIELYLPSEGVYIKNNMADFTGEADPOGNGENTPSENGK-VSTGTV 936
 DB 1516 AHVAKITSLDKALIDKIDLF-----KNTNDFEAIKRLINDTFRKMDKGLISTGLV 1567
 QY 937 ENQPT-----ENKPADSLPEAPNE--KPKPENS 963
 DB 1568 QNFPTIISKLEGGFQDMINIQQHOCVKKQCPENS 1603

RESULT 3
 ID CNA STRAU STANDARD; PRT; 1183 AA.
 AC 053654;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE COLLAGEN ADHESIN PRECURSOR.
 GN CNA.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Staphylococcus.
 OX NCBI_TaxID-1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FDA 574;
 RX MEDLINE-92165839; PubMed-1311320;
 RA Patti J.M., Jonsson H., Guss B., Switala L.M., Wlberg K.,
 RT "Molecular characterization and expression of a gene encoding a
 Staphylococcus aureus collagen adhesin."
 RL J. Biol. Chem. 267:4766-4772(1992)
 RN [2]
 RP ERRATUM
 RA Patti J.M., Jonsson H., Guss B., Switala L.M., Wlberg K.,
 RL J. Biol. Chem. 269:11672-11672(1994).
 RN [3]
 RP COLLAGEN-BINDING DOMAIN.
 RC STRAIN-FDA 574;
 RX MEDLINE-94032261; PubMed-8218209;
 RA Patti J.M., Boles J.O., Hoeck M.,
 RT "Identification and biochemical characterization of the ligand
 binding domain of the collagen adhesin from Staphylococcus aureus."
 RL Biochemistry 32:11428-11435(1993).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
 RX MEDLINE-97475225; PubMed-9334749;
 RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
 RA Moore D., Jin L., Schneider A., Delucas L.J., Hoeck M.,

RA Narayana S.V.L.;
 RT Structure of the collagen-binding domain from a Staphylococcus
 aureus adhesin.
 RL Net. Struct. Biol. 4:833-838(1997).
 CC -1- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO
 COLLAGEN-CONTAINING SUBSTRATA.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN, CELL WALL.
 CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
 IN THE REGION OF THE MEMBRANE ANCHOR.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M81736; AAA20874.1;
 CC PDB: 1AMJ; 24-JUN-98.
 DR InterPro: IPR001899;
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
 DR Signal: Repeat; Transmembrane; Cell wall; 3D-structure.
 FT SIGNAL 1 29
 FT CHAIN 30 1183
 FT DOMAIN 1158 1177
 FT DOMAIN 1178 1183
 FT DOMAIN 151 318
 FT DOMAIN 533 1093
 FT DOMAIN 1093 1157
 FT DOMAIN 1151 1156
 FT REPEAT 533 719
 FT REPEAT 720 906
 FT REPEAT 907 1093
 FT SEQUENCE 1183 AA; 133066 MW; B6A1CC072E575D76 CRC64;
 Query Match 3.88; Score 205.5; DB 1; Length 1183;
 Best Local Similarity 19.78; Pred. No. 0.03;
 Matches 214; Conservative 139; Mismatches 411; Indels 323; Gaps 50;
 QY 36 NNRVSYVDGSSQSKSENLTR-PQGVSOKEGIAEQIYIKITDQGYTSHGHYHYNGK 93
 DB 195 NNEKSYV-----SNDITIKQIQQGGQQLDSTLNINVT-----GHSNYTSGS 236
 QY 94 --VPDALFSEELMKDPNTQDKADI-----VNEVKGYYIIRVDG---KYVYLLDAA 142
 DB 239 AITDFEKAFPSKLTIVNTKNTIDVTIPQGYGSYNSFSINYKRTITNEQKEFVNSQAW 298
 QY 143 HADVNRKDEINROKOEHYDNENVSNNV-----AAVRSQGRITTN 183
 DB 299 YOEH--GKEEVNGSFSNHTVNIINANGISCTYVGEIKLVKQDKTRAPLANVFKLSKR 356
 QY 184 DGYVF--NPADITEDTGNAVYIPHGCHYIYIKSDLSASELAAKAHLAGKAMPQSOL 239
 DB 357 DGSVYKDNQKREILITDANGI-----ANKALPESGYIILKEIAPRPYFDKAE----- 406
 QY 240 SYSTSTASDNNTQSYAKSGTSKRPANKSENILQSLKELDSPSAQRYSSDGLVDFDPATIS 299
 DB 407 -YPTMTDJDQGYF--TTIENAKAIKTKDY-----SAQR-----VNEGTDKVK 448
 QY 300 RTPGVAVIPRODHHYIPYK--LSALEKIAKRVPISGGSYSTVNAKNEVYSSLSGL 357
 DB 449 PTYFKLYKDDNONTTPVKAELKLEED-----GTTKYWSMLPND----- 491
 QY 358 SSNPSLTTSKELSSADGYIFNPKDIVEETAAAYIVRHGDHFIYKRS--NOIGOPTLP- 415
 DB 492 -----KNKAKIKYLVKEVNAAGEPTTPE 514
 QY 416 -----NNSLATPSPSLDI-NGTSHEKHEDDGFGFDANRIIADESGFVNSHGDHNYTF 469

DB 515 GYTRKENGVLVNTNEKPLETTISIGERWMDKDNQDKR--PEKVSYNLLANG----- 565
 QY 470 KKDLEBQIKAAQHLEEVKTSHGGLDLSISHEDYDGNAMKEMLDKITE-EKIAIGMK 528
 DB 566 -----EKVKT-----LDVTSNNKT-----EFKDLKRYDGGKIEYTV 600
 QY 529 QYVK-----RESIYVNR-----EKNAIYPGHHHADPIDERPYGIGHSHSNYELFK 578
 DB 601 EDHKDVTDTINGTTINKTTPGTSATVTRKMDNNQ--DGRPPEI-----KVLEY- 652
 QY 579 PEEVA-----KKEGNTV-YTGEELIVNVLLKSTNNQNT 615
 DB 653 -QDKAKGKTALINESNNWHTWTGLDEKAKGQQVKTVEELFVY-----KGYTHVND 707
 QY 616 LANGKRVSPPELKKIGINLVLLIPDGVLEKSVKVGSG-----VGNIAN 668
 DB 708 M--GNLIVTKNTYPTETTSISGERWMDKDNQDKRPERVSNLADDEKKTLDVTSNT 765
 QY 669 FELDQYLP-----GQTFKYTAS--KDYPEVSDGTFTVPTSLAYKASQTIFFYPHAG 721
 DB 766 WKYEFKLPKRYDESKKLEYTVTEHDVNDY--TTDINGT-----TINKKTPG 810
 QY 722 DTYLRV-----NQRVVPKGTALVYFDE-----FHGNAYLENNYKV 759
 DB 811 ETSATVTRKMDNNQDKRPTKEIKVELYQDKATGTALINESNNWHTWTGLDEKAKG 870
 QY 760 GEILPPIKLNQ--GTRFAGNKIPVTPANAVLNDQSYIYVPILEKENQTDK---- 812
 DB 871 QQVATYIEELKRYKGYTHYDNDNGMLIYNTKTPETTSISGERWMDKDNQDKRPER 930
 QY 813 --PSILPOFRNKA-----QENSKLDE--KVEPKTSEKVEKESLSENGST 855
 DB 931 VSVVLLANGKRVKTDVTSNNKTEFEKDLPKYDEGKIEYTVTEHDVNDYTTDINGTTI 990
 QY 856 SNGSL--EVPPTYD-----PVQRYAKFSESGMKLENTLVNNDGTYIELYLP 900
 DB 991 TNYTPGTSATVTRKMDNNQDKRPTKEIKVELYQDKATGTALINESNNWHTWTG 1050
 QY 901 SGEYIKKNMADFT--GEAPDNGENKPSSENGKRVSGVTEQNTPEKRPADSL--PEAPNKR- 956
 DB 1051 LDEKAKQGYKYVDELTKNYGTTHYDNDNGMLIYNTKTPKPKPKPIYPEKPKDTP 1110
 QY 957 PVKPENS 963
 DB 1111 PTPRDS 1117
 RESULT 4
 MSPL_PLAFC STANDARD; PRT; 1726 AA.
 AC P04934
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MERZOITE SURFACE PROTEIN 1 PRECURSOR (MERZOITE SURFACE ANTIGENS)
 DE (PMSA) (P195).
 GN MSP-1
 OS Plasmodium falciparum (Isolate Camp / Malaysia).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID:5835;
 RN (1)
 RP SEQUENCE OF 1-1103 FROM N.A.
 RX MEDLINE-86205236; PubMed-3517809;
 RA Weber J.L., Leininger W.M., Lyon J.A.;
 RT Variation in the gene encoding a major merozoite surface antigen of
 the human malaria parasite Plasmodium falciparum.
 RL Nucleic Acids Res. 14:3311-3323(1986).
 RN (2)
 RP SEQUENCE OF 1104-1726 FROM N.A.
 RX MEDLINE-88143999; PubMed-3278296;
 RA Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;
 RT Merozoite surface protein sequence from the Camp strain of the human
 malaria parasite Plasmodium falciparum.

RL Nucleic Acids Res. 16:1206-1206(1988).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POTENTIAL)
 CC -1- PFM: MEMOZITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEMOZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: X03831; CA27446.1;
 DR A23386; SAQGM.
 DR InterPro: IPR000561;
 DR Pfam: PFO0008; EGF, 1.
 KW Malaria; Memozite; Polypeptide; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL: MEMOZITE SURFACE PROTEIN 1.
 FT CHAIN 20 1726
 FT CARBOHYD 133 133 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 272 272 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 501 501 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 567 567 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 638 638 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 827 827 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 924 924 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 944 944 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 990 990 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC...) (POTENTIAL)
 FT SEQUENCE 1726 AA; 196197 MW; DDA045FA352BCF3 CRC64;

Query Match 3.88; Score 205.5; DB 1; Length 1726;
 Best Local Similarity 20.08; Pred. No. 0.051;
 Matches 246; Conservative 172; Mismatches 414; Indels 397; Gaps 65;

1 MKFSKTYIAGSAVYISLCAVYALNDRSEN---KDNRYVYVDSOSSOSSENITP 56
 531 MKFNNE---DKDYVDKIFSAHYTYNEKORYNKFSSNSVYV---QKLKALSTLE 584
 57 DOVQKESIGAEQVYIKITDGYVTSQDHYHYNGKVPDA---LFESEI----- 104
 585 D-YSLRKGI-----SEDFNHYYTLKLGLEADIKKLTLEIKSENKILE 627
 105 ---LMDPYOLKADIVN-EVKGYYII-KVDS-KYVYVLDAAHADNVRT----- 149
 638 KNFKGLTSHANASLEVYDIYVLOQVYLLIKKIEDLKIELEFKNAQKDSIHVPNIYKP 687
 150 ---KDEINRKO-----EYVQDNKRYNSVNA---VASQGYTTRDGYVF 188
 688 QNRPEPYLYIVKKEVDLKEPIPVKMDLKEQVYLSISIPQVPAASE---TTEDS--- 741
 169 NPADIEDTGNAYIVPHGHHYIPKS-----DLASELAAKAHLAAGNMOPSOISY 241
 742 ---GHSHTLISQSETEVETETETETETETETETETETETETETETETETETET 784
 242 SSTA---SDNTQSVAGSTSKPANR-----SENIQSLKELYD-SPSA 281
 745 VENSLEHNSNDNSQALTYVYIAKIDELFTSYCHKIIVSNSMODKLEVVYVLPFEE 844
 283 QRTSSDGLVEDPAKIIISRTNGVAIPRG-----DHYHEIPYSK-----LSALEEK 327
 845 ENELKS---CDPLDLNLNIONNIPAMYSLYDSNNNDLQHFFELYKREMYIYKLEKEE 900
 328 ---IARMPISGTGTVST-NAPKPEVYVSLGSLSNSPSLSTSELSASGCIYFN 380

901 NHKKLEEKQITGSSSTSPGTYVNTAASATSHNSQOQNAS---TNGONG----- 953
 381 KDIVETAAVAYVVRHGDHPHPIKSNQI-GQPLPN-NSLATPSPSLPINFQSHK 437
 954 ---VAVSSGAVVEESHDPLTVISNDKGIYSLNMGKTKVPPN-LTIST-TEMEKP 1008
 438 EEOGYGFDAARITAEDESGVMSHGHNNHFFPKD-----LREQKAKQKLEHYKTS 491
 1009 YEN-----ILKNNDYTP---NDIKQFVNSKRYTGLTETQNNALNDEKTKKDT 1056
 492 ---HN-----GLDLSHEDDYPGNAMEND---LDKRIEKIAGIKQGVKRE-S 536
 1057 LQISFDLYNNYKIKLIDLEFNKKELGODKQKQIKKLTLEKQESKLSLNNPHVLONS 1116
 537 IYVKKENALII-----YPHGDHHRADPIDERKPVGI-GSHS 572
 1117 VFFPKKKEAEIENLENTLENTKILKKYKGVYKYNE---SSPLKTLSEVSIQEDNYA 1173
 573 NYELFKPEGVAKKRGKNTYTGEE-----LTNVNVLKNSIFNNQNTLNGQKRV 624
 1174 NLEKFRVLSIDGKLNNDNLHGKRLKSLFSGHLHLLTEIK-EYIKRKNYI----- 1223
 625 FSPPELEKLGIMLVYL---LTPDGYLEKVGSGVGNIANFELDOPYLPQOTF 681
 1224 GNSPSENNKR--VNEALKSTENLPEAKVTVVTP-----PQDVTSP 1266
 682 KYTI-----ASKDPEVSYDGTFTVPTSLAYKMASQTFYFPHAGDYLYRVQFAVPKG 736
 1267 SVRVSSGSGSTKEHTQIPTSG-----SLTEIQVVOVLOANYDEEDSLVLPFEGSED 1320
 737 TD-----ALRVFDEFHGAIVYENNYKGEIKLPKLNQGTTRRANKIP---V 783
 1321 NDELDQVTVGEAISVTMDNLISG--FENEVDYILK-PL---AGYRSIKKQIEKNIF 1373
 784 TPANA-----YLD-----NOSTYIVYVPILEKNOTDKPSILPQ 818
 1374 TFINLNDIINSRLKRRKRYFLDVLSDLMQFKHISSEYITE-DSFLLNSQKNTLKS 1432
 819 FRRNR-----AOENSKIDEKV-----EPTSEKEREKLSGTGNSSTNLE 861
 1433 KYIKESVENDIRKQAGISITRYEVYAKYKDDLESIKYIKEREKRPSPPTTPPS--- 1489
 862 EVPTVDVQEVKAFNAFA-----EYGNKLENTVLENM----- 891
 1490 --PAKTDEQKESKFLPFLNIEFLYNNLVKIKIDDYILNKAINDCQVEDEAHVITK 1547
 892 ---DGTIELPISGEVIAKKNADFTGAPGNGENBPSENGK-VSTGYVNOPT-- 941
 1548 LSDKAIIDKIDLE-----KNHNDFEAIKRLINDYKMDKGLSLSTGLVONFPRTI 1599
 942 ---ENKPADSLPEAPNE-KYKPPNS 963
 1600 ISKLIIEGFQDMINISQHCYKQCPNS 1628

RESULT 5
 MSP1_PLAAP STANDARD; PRT: 1726 AA.
 AC P50495;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MEMOZITE SURFACE PROTEIN 1 PRECURSOR (MEMOZITE SURFACE ANTIGENS)
 DE (PMSA) (GP195).
 GN MSP-1.
 OS Plasmodium falciparum (Isolate Palo Alto / Uganda).
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 OX NCBI_TaxID=51270;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=99005525; Pubmed=3049134;
 RA Chang S.P., Kramer R.J., Yamaga K.M., Rato A., Case S.E.,
 Siddiqui W.A.;

"Plasmodium falciparum: gene structure and hydrophobic profile of the major merozoite surface antigen (gp195) of the Uganda-Palo Alto isolate." Exp. Parasitol. 67:1-11(1988).

- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (POTENTIAL).

- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42 KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

CC This SWISS-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: M37213; AAA29611.1; -

DR InterPro: IPR000561; -

DR Pfam: PF00008; EGF_1

KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor

FT SIGNAL 1 19

FT CHAIN 20 1726

FT CAROHTD 133 1726 MEROZOITE SURFACE PROTEIN 1. (POTENTIAL)

FT CAROHTD 272 1726 N-LINKED (GLCNAc) (POTENTIAL)

FT CAROHTD 501 1726 N-LINKED (GLCNAc) (POTENTIAL)

FT CAROHTD 567 1726 N-LINKED (GLCNAc) (POTENTIAL)

FT CAROHTD 638 1726 N-LINKED (GLCNAc) (POTENTIAL)

FT CAROHTD 827 1726 N-LINKED (GLCNAc) (POTENTIAL)

FT CAROHTD 924 1726 N-LINKED (GLCNAc) (POTENTIAL)

FT CAROHTD 944 1726 N-LINKED (GLCNAc) (POTENTIAL)

FT CAROHTD 990 1726 N-LINKED (GLCNAc) (POTENTIAL)

FT CAROHTD 1016 1726 N-LINKED (GLCNAc) (POTENTIAL)

FT CAROHTD 1114 1726 N-LINKED (GLCNAc) (POTENTIAL)

FT CAROHTD 1221 1726 N-LINKED (GLCNAc) (POTENTIAL)

FT CAROHTD 1613 1726 N-LINKED (GLCNAc) (POTENTIAL)

FT SEQUENCE 1726 AA; 196174 MW; 5B59CEFA29A026 CRC64;

Query Match 3.8%; Score 205.5; DB 1; Length 1726;
Best Local Similarity 19.7%; Pred. No. 0.051;
Matches 243; Conservative 172; Mismatches 413; Indels 405; Gaps 64;

QY 1 MFSKRYIAGSAVIVSLICAYALNOHRSQEN---KDNHVSVDGSSQSKSENLP 56
DB 531 MFSNNF---DKDYDKIFSAHYTNVEKQRYNKFSSNNNVYV---QAKAKALSTLE 584

QY 57 DOVSOREGIAEQIYIKITDQGYTSHGDHYHYNGKVPYDA---LFSEEL----- 104
DB 585 D-YSLRKGJ-----SEKDFNHYYTLKGLLEADIKKLFEBIKSSEKILE 627

QY 105 -----LMDPNYQLDADIVN-EYKGYII-KVDG-KYYVYLKDAHADNRT----- 149
DB 628 KNEKGLTHSANSLSHYIVDKLVOKVLLIKRIEILKLEFLKNAOKLSDIHVNRYKP 687

QY 150 -----KDEINRQK-----EYKDNKFNYSNVA---VARSGRYTNDGYE 188
DB 688 QNKPEPYIIVLAKREVDKLEKIPRYKDKMKKEQAVLSITQPLVASE---TTEDG--- 741

QY 189 NPADIIETGNAYIPHGHHYIDKS-----DLASSELAAFAHLAKGNMOPSLST 241
DB 742 -----CHSTHTLSQSETEVEETEELVEGVGHTTYITLTPREKVV 784

QY 242 SSTA-----SDNNTOSVANGSISKPANK-----SEUOSLAKELYP-SFSK 281
DB 785 VENSLEHKSNDNSQALTYIVIKLDELTSYCHKILVSNSSMDKLEEVNWTPEE 844

QY 282 QRTSSDILVEPARIIRTPNGVAIPHG-----DHYHFIYSK-----LSALEK 327
DB 845 ENEELKS-----GDPDLILNIGNNIPAMSLVDSMNDLQHLFEIYQKEMITYLKLEE 900

QY 328 -----IARVPLSGTGVST-MAPNEVSSGLSSPSSLTTSKELSSASDQITFN 380
DB 901 NHIKKLEKQKITGSSSTSSPNTVYVYVQSAHSSNQSQSNASS---TNTQNG----- 953

QY 381 PROIYETAAIVRGRDHYHYPKSNQI-GQTLDPN-NSLATPSPSLPINCSTHERH 437
DB 954 ---VAVSGFAVEESHDPILVISINDKIVSLNKGKIVPVP-LTIST-TENEH 1008

QY 438 EEDGYFDANRIAEDESGVSHGDHNYFFKRD-----LSEQIKAAOKHLEVYTS 491
DB 1009 YEN-----LAKNDYTF---NDIQFVNSKRVITGLTEPKNALNDIKKILFD 1056

QY 492 -----HN-----GDSLSSEHDYFGNAKEMD---LDKRIEKIAGIMQYGVARE-S 536
DB 1057 LQSLFDLYNKKILDLRNFNKKELGQDKMQIKKILKLEOKSLKSLNPNHVNQNTS 1116

QY 537 IYVNEKNALI-----YRPGDHHDAPIDENKPVGI-GHSHS 572
DB 1117 VEFNKKREKIAETENTLNTLNTLKHRYGLVRYNGE---SSPLTLEVSISTEDNYA 1173

QY 573 NYELFPEEGVAKKGNRYTGE-----LTVVNLKSTNNONFTLANGQ 620
DB 1174 NLEKPVLSKIDGKLDNHLKKSIFSSGLHOLITELKEVYKKNYTGNS----- 1226

QY 621 KRVSSFPPELEKRLGINNLVRL---TPDGKLVKFSGVGEVGNIANPELDQPLP 677
DB 1227 -----PENNNK---VNEALKSYENFLPEAKTYVTE-----PGDVT 1262

QY 678 GQTEKTYI---ASNDYEVSYDGFYPTSLATMASQITFPFHADDTLRVAPORA 732
DB 1263 PPSLVKRVSSGSSSTEEQIPTS-----SLTELQOVVOLQWDEDDSLVYLPIPG 1316

QY 733 VPKGTD-----ALVRVDEFHGNALYENKRVGEIKLPRLNCGTTRTAGNKIP 782
DB 1317 EEDNDVETDQVYGEAIVYMDNLISG---FENEIDVYILK-PL---AGVRSLSKQLE 1369

QY 783 ---VTFMANA-----YLD-----NOSTYIVEVPILEKENDTKPS 814
DB 1370 KNIEFNLINDILNSRLKRRYFLDYESDLMOFKHISNRYIE-DSFKILNSQKNT 1428

QY 815 ILDFKRNK-----AQENSKIDRY-----EPRKSEVEREKLSGTGNSFN 857
DB 1429 LKSTKYIKESYENDIKRAQEGISYTERVLAHYKODLESIKRYIIEKEKFPSSPTPP 1488

QY 858 STLEVPYVDPOEKAKFA-----ESYGMKLEVNLFNM----- 891
DB 1489 S-----PAKTQKESKFLPLTNIETILYNNLVAKIDYILNLAKINDCAYEDEAHV 1543

QY 892 -----DGTIELYLPSEVYIKKMADETFGEAPQNGENKPSNGK-VSTGVENO 939
DB 1544 KITKLSDKAIDKIDLF-----KNHNDDAIKKILINDTKDMGKLSLGLVQNF 1595

QY 940 PT-----ENKPADSLPEAPNE-KPVKPNSS 963
DB 1596 PTIISKLEIGKFDMLNISQHCYKQKOPENS 1628

RESULT 6
MRSP_STAUD STANDARD; PRF; 1637 AA.
ID MRSP_STAUD STANDARD; PRF; 1637 AA.
AC P80544; 09Zf62;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MERICILLIN-RESISTANT SURFACE PROTEIN PRECURSOR
GN PLS.
OS Staphylococcus aureus.
CC Bacteria; Firmicutes; Bacillus/Clostridium group;
CC Bacillus/Staphylococcus group; Staphylococcus
OX NCBI_TaxId=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ISOLATE 1061;

RA Savolainen K., Kussela P., Paulin L., Korhonen T.K.:
 "Pis, a large repeat-rich surface protein of methicillin resistant
 Staphylococcus aureus." submitted (Dec-1998) to the EMBL/GenBank/DBJ databases.
 [2]
 RC STRAIN-ISOLATE 1061;
 RX MEDLINE-96670743; PubMed-8665912;
 RA Hilden P., Savolainen K., Tynnelae J., Vuento M., Kussela P.;
 "Purification and characterization of a plasmid-sensitive surface
 protein of Staphylococcus aureus."
 RT Eur. J. Biochem. 236:904-910(1996).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AF15379; AAD09131.1;
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR PROSITE: P000343; GRAM_POS_ANCHORING; FALSE_NEG.
 KW Cell wall; Transmembrane; Antibiotic resistance; Glycoprotein; Repeat;
 KW Signal.
 FT SIGNAL 1 48
 FT CHAIN 49 1637 METHICILLIN-RESISTANT SURFACE PROTEIN.
 FT DOMAIN 1301 1582 141 X 2 AA TANDEM REPEATS OF D-(SAG).
 FT DOMAIN 1598 1603 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
 FT PROTEINS.
 SO SEQUENCE 1637 AA; 174573 MW; 75BE9ADB46BD309 CRC64;

Query Match 3.8%; Score 203.5; DB 1; Length 1637;
 Best local similarity 19.0%; Pred. No. 0.058;
 Matches 228; Conservative 144; Mismatches 462; Indels 365; Gaps 50;

QY 2 KRSKRTIAAGSA-VIYSLSLC-----AYALNQRSGENKNNRKYVDGS-----QS 47
 DB 18 KRSIRRTVGTASILVGAFLIFGVANDQAEAEENNTTKODSSDSKVGANVTIOGSS 77
 QY 48 SQRSEMLPDDOYQKSGIOAEQIVIKITDOGYVTSBGDHYHNGRYVPDALFSEELMK 107
 DB 78 ANSNESDIPEQVD-----VTKDTE-----QASTEEKANTT 108

QY 108 DPNYQDKADIYEVNGYIIKVDGYYYLKDAAHADNVKTEINRQKOEHVKNERY 167
 DB 109 EQASTEERKADTTEQATEEAPRAEAGTGDKYTEEAPRAEET--DKATTEEAPRAEETDKA 165
 QY 168 NSNVAVARSGQRTTNDGYVFPNADIIEDTGNAIYVPHGHHYIKRSLD-----218
 DB 166 TEERAPTEETDKATTEE-----APAAETSKAT-----EAAKAEETSAATEEAP 212
 QY 219-----ASELAANKAHLAGKMPOLSYSTASDNN 249
 DB 213 KAEETKATEEAPKTEEDDKYTEEAPRAEETSKAATEAPRAEETNKVETEAPAAE 272
 QY 250 TQSVANGSTSPANKSENIQSILK-----ELYDSPAQRYSSSDGLVDPANIIIS 300
 DB 273 TNKAAEET--PAVEDTNKSNNAOPSETERQVYDVAKDLYKKSEVTEAKAELEKY 330
 QY 301 TPNGVAILPHGDHYHFTPYSLALE-EKIRAMPISGTGVTNKPKNVYSSLSLS 359
 DB 331 LPMDI-----SMLSNEETIKKIALSEVLEKETAN--KENAPRATFRSVSSNAR 375
 QY 360 NPESLITKSLASADGYIENPDYIEETATAVIRHGDHFHPIKPSNQIGOPTL-----414
 DB 376 TTVNNTSATLRAAA-----ODVYTKGTGFTLAGDIIHTTYKEEPNESTILAFNT 438
 QY 415 ---PNNSIATP-----SPSLINPGTSHEKHEEDGYGPDANRIIAD--ES 455
 DB 429 NFNPNQTKKALRYNDKIDENKDFLITVYVANNNGQNTTADQMGVFTFGNGQODFLNG 488

QY 456 GFVMSGHGDNHYFFKKDLTEQIKAAQKHLSEYKTSHGSLDLSHEDDYPGNAAEMDL 515
 DB 489 GILRQKGNANSGFKIDPAAVNVNKGKVDLADKT--NLSIQIGAKAGYGTFFVNGAD- 545
 QY 516 DKITEKTAGIKQYGVKRESIVYKER--NAIITPHGHADPIDDKPPVIGSHS- 572
 DB 546 -----GVNNOVG-----QNALNTRKDPVNNIITADNTTNLD-----GQFHQ 583
 QY 573 -----NYELKPRGVAKKRGKRYTGEELTNVNNLKNSTFNNQNTLANQKRRVSF 625
 DB 584 RINDVYLAND-----AATSIITATYAGTKATKATTDGIDSKQYNFLIISSHQNY 636
 QY 636 S---PPELEKKGIMMLVKLTTPDQKYLEKVSAGVSGVGNANFELDQYLPQGTQK 682
 DB 637 SNGIRTNLEG-----VTITTPQADLIDV--ETKQPIPKITREEDPILERS--- 684
 QY 683 YTIASNDPVEV---SYDS--FTYPTSLATYKASGTYIPPHAGDTYLRVNDQFVPG 736
 DB 685 -----PDVIVQKGEDEKTTTPT---KVDPT-----GDVERGEPTTEVTN 725
 QY 737 -TDALVRY-----DEFGNAYLENNYKGEIKLPKLNQSTYTAGNKIPIVTEM 786
 DB 726 PVDEIVHFTPEEYVPOGHKDEFPN-----LPI-----DQTEVPG----- 760
 QY 787 ANAYLDNSTYIVVPILEK-----ENQDKPSILPQKRNKAQENSKIDEREVP 837
 DB 761 -KPGIKNPETGVVPPVDVYTKHGPKAGEPEVTEKIEIFEKREPNPLKGEERY--- 816
 QY 838 KTSERVENKELSETGNSNSLSLEY---PVDPQOEVAFAEYQKLE-----N 886
 DB 817 -TOEGQGTKEKTTPTTPTINPLGEEKVGEPEPTTEVKEPVDELTOGGEVPOGHKDEPD 875
 QY 887 VLENNQGTIELVLP-----SGEVIKNNADFTGEAPQGN----- 920
 DB 876 PMLPDTGTEE--VPGKPGIKNPETGVVPPVDVYTKHGPKAGEPEVTEKIEIFEKREPF 933
 QY 921 -----GENKPSBNGRVSTGV-----ENQPT---ENKPADSLPEAPNPK 956
 DB 934 NDLKPKGEKTVQEGQGTGKTTTPTTINPLGEEKVGEPEPTTEVKEPVDELTOGGE 993
 QY 957 -----PAKPNSTDN-----GNLMPK--GNVSSDMLDPALEAPVNDVQOKLE 999
 DB 994 VPOGHDEPDPMLPIDGTEVPKPKIKNPGEVVTVPVDVYTKHGPKAGEPEVTEK 1052

RESULT 7
 ID IG44_HAEIN STANDARD; PRT; 1849 AA.
 AC P45386;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE IMUNOGLOBULIN A1 PROTEASE PRECURSOR (EC 3.4.21.72) (IGAL PROTEASE).
 GN IGA.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=NH1 HK61;
 RX MEDLINE-92234949; PubMed-1373717;
 RA Poulsen K., Reinholdt J., Killian M.;
 "A comparative genetic study of serologically distinct Haemophilus
 influenzae type 1 immunoglobulin A1 proteases."
 RT J. Bacteriol. 174:2913-2921(1992).
 RL FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMUNOGLOBULIN A
 CC PRODUCING INTACT FC AND FAB FRAGMENTS.
 CC CATALYTIC ACTIVITY: CLEAVAGE OF IMUNOGLOBULIN A MOLECULES AT
 CC CERTAIN PRO-XAA BONDS IN THE HINGE REGION. NO SMALL MOLECULE
 CC SUBSTRATES ARE KNOWN.
 CC -1- SUBCELLULAR LOCATION: SECRETED.

CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
 CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
 CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
 CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M87491; AAA24968.1; -
 CC DR MEROPS: S06.001; -
 CC DR InterPro: IPR000710; -
 CC DR PRINTS: PRO0921; IGASERPRASE.
 CC KW Hydrolyase; Serine protease; Transmembrane; Zymogen; Signal.
 CC FT CHAIN 1 25 POTENTIAL.
 CC FT PROPEP 26 1021 IMMUNOGLOBULIN A1 PROTEASE.
 CC FT ACT SITE 1022 1849 HELPER PEPTIDE (POTENTIAL).
 CC FT ACT SITE 299 299 PROBABLE.
 CC SO SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;
 query Match 3.8%; Score 203.5; DB 1; Length 1849;
 Best local Similarity 19.3%; Pred. No. 0.069;
 Matches 229; Conservative 153; Mismatches 449; Indels 397; Gaps 59;
 QY 28 HQEKKNNRRSYVDGSSQSKSENLTPDQVSQKEGIG- AEOIVKITDQ 77
 DB 125 HRVSEENR---YTVENKNEPTENVSTFKKEEDDAKRREDDYMPRLDKVTVAVRI 181
 QY 78 GYTSGDHYHYNGKVPDALF---SEELMDPNOL---RDAD---YNEVRKG 125
 DB 182 EASTANNKE-YNSDKTPATPVRLGSGQFYTKGSRQLITTEKOGNLLRMNDVG 240
 QY 126 YIIRKDGKYYVYLKDAHADNVRTKDEINRKOEHYKDEKYNVAVARSGRYTTNDG 185
 DB 241 DNEIYGNMYT---GIAGTPYKVNEN---NGLIGPNSKEEHSDDPG 283
 QY 186 YV---FNPADIIETGNAIVY---PHGHIYTPKSDL---SASLAARAKHLAK 232
 DB 284 ILSODELTNYAVLGSGSLFYDREKGMFLGSDYFAGYKNSQWQMNTYKHEFAEK 343
 QY 233 NNQPSOLSYSS-TASDNTQSV--AKGSRKPAKSENGSLKELYPDSPAKRY----- 284
 DB 344 IYQ---QTSAGSLGSKNTQYTWQATGSGYSTTGGGEPLSDVLTGCKDPNPKGSITLKG 399
 QY 285-----SESDQLYVDPARKIISRTPN-----GVAIDPGD---HYHFIPTSKLS 322
 DB 400 SGTLLNNHIDGAGLGEFEGDYEVKGTSDSTTWGAGVSADGKTYWTKVNPRTDRA 459
 DB 323-----ALEEK-----IARMPISGTGSTVSTNPK- 346
 DB 460 KIGKGLVVEGKNEGELKLVGDGYVILKOKADANNKYQAFSGVIGSRSTLYLNDKRO 519
 QY 347--PNEVYSSL--GSLSNPSSL-----TTSKELSSASGCTFENPKD 383
 DB 520 VPPNSIYRFGRGRDLNAGNSLTFPHIRINIDGAAVNNHNMNTNITITGSLTNP- 577
 QY 384 IVEETATATVYRHGDHFH-----YTPKSNQIGOPTLPNNSLATSPSPILPBG 431
 DB 578---NNTTSYNIBADDDHPLRISIPYQVLYFNQDNR--SYTLKKG--ASTRSELPQMSG 631
 QY 432 TSEHNEHEGIGYFD-----ANRTIAEDSGGVVSHGHNHYFFKKDLTEQIKAAQKHL 485
 DB 632 ESNEMWLYMGRTSDAKRNVNHNINNEHNGF-----NGYF-----GEEETKATON- 677
 QY 486 EYVTSKSHGLDLSLSEHODY--PGNAKEMKDLDKKIEKINGIMKQYGVKRESYIVNKK 543

DB 678 GKINVTENG---KSDQNRRLTGNTLNGDL-----NVER 709
 QY 544 NAIYVGHGHHADPIDEKPVGIG-----HSHSYTELEKPEGVAKK-----EGN 589
 DB 710 GTFLGGRPPPHARDI-----AGISSTKDPHTTENVEVEDWNNRNFATNTANTGN 764
 QY 590 -KYVTEBELTNVNLKNSFNQNTFLANGQKRVSEFPELEKIGIMLVK-----642
 DB 765 ASLYSGR--NVANITSNITAS--NAOHIGY-----KTGDVYCVRSYTG 807
 QY 643 LITPDQVLEKSGKYFGESV-GNT-----ANFELDQVILPGQTFYTTIASKDYPEVSD 656
 DB 808 VYCHNSNLSERKALNSFNPTLRNVNLTEVASPTLGKNIIFG-----TQISIGTSQVN-- 860
 QY 697 GFTVPTSLAYKASQITTFPHASDTYLRVNOQFAVPRGTALVAVPDEFHNA-----751
 DB 861---LKNHSHWHLTGNSNVNQLTNGHHLNQNQDANKVYTNLTITVNSLGSNGSFYTW 916
 QY 752-----YLENNKYVEIKLPI-----PLNNGT-----TRTAGNKIPVTFMANAYL 791
 DB 917 YDETNRKSNKYVYKNSATGNFTLVQADKTGEPHNNELTFLDASNATRNMLEYV-LANGSY 975
 QY 792 DN-----QSTYIVEPVLEKENOT-----DKPSILPQFRNRQAQEN 827
 DB 976 DRGAMRYKLRNVNGRDLYNPVEYKRNQTVDTNITTPNDIQADAPS-----AQGN 1026
 QY 828 SKIDKEVEEP-----KTSRKVEKELSENGST 855
 DB 1027 NEIARVETVPVPPAPATSAIASQEPETRPAPETAPAEETNTASTTAPSDATQOT 1086
 QY 856 SNSTEEVPTVPDVEKVAKFAESYGMKLENYLFNNDGTIELYLSGSEYIKRMADF-- 912
 DB 1087 ENPNSBSVPS-----ETTEKVAENPQENETVANKNOEATEPTLPNGEVAKADQVTEAN 1141
 QY 913--TGAPQNGENKNSKSENGKSTGYVENOT-----EKKPADSLPEADPEKPY-----K 959
 DB 1142 TQTNARATSEGTETETO-----TAEKSEPTESYVYSENOPEKTVYOSATEDYVVEKEEK 1196
 QY 960 PENSTDNGMLNPEGVNGSDP-MLDPALAEAPVADPVOE 996
 DB 1197 AVETIEEYOKAPQVTSKEPKQAEPAPEVPIDTNAEE 1234
 RESULT 8
 MSP1_PLAF3 STANDARD; PRT; 1682 AA.
 AC P19598; Q25921;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
 DE (P190).
 GN MSP-1.
 OS Plasmodium falciparum (isolate ro-33 / Ghana).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID:5834;
 RN [1]
 RP SEQUENCE OF 1-1061 FROM N.A.
 RX MEDLINE-8816657; PubMed-3327688;
 RA Cetta U., Rotmann D., Matile H., Reber-Liske R.,
 RT "A naturally occurring gene encoding the major surface antigen
 RT precursor p190 of Plasmodium falciparum lacks tripeptide repeats."
 RL EMBO J. 6:4137-4142(1987).
 RN [2]
 RP SEQUENCE OF 1032-1682 FROM N.A.
 RX MEDLINE-95354793; PubMed-7628566;
 RA Tolle R., Bujard R., Cooper J.A.,
 RT "Plasmodium falciparum: variations within the C-terminal region of
 RT merozoite surface antigen-1."
 RL Exp. Parasitol. 81:47-54(1995).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POTENTIAL).
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42

KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
MERIZOTES. THE MATURATION TAKE PLACE DURING SCHIZONT.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: M35727; AAA29715.1;
DR EMBL: Y00087; CA868280.1;
PIR: S06286; S06286.
InterPro: IPR000561;
DR Pfam: PF00008; EGF, 1;
KW Malaria; Merizotes; Polypore; Repeat; signal; glycoprotein;
FT Transmembrane; GPI-anchor.

FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1682 MEMBRANE ANCHOR (POTENTIAL).
FT TRANSMEM 1666 1682
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 785 785 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 881 881 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 901 901 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 947 947 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1178 1178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1569 1569 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1682 AA; 192462 MW; C82A1E15948CADC6 CRC64;

Query Match 3.5%; Score 189; DB 1; Length 1682;

Best Local Similarity 20.0%; Pred. NO. 0.28; Mismatches 428; Indels 362; Gaps 64;
Matches 242; Conservative 177;

QY 1 MKRSKRYIAAGSAIVISLCAAYALNOHRSQEN---KDNRRVSYVDSGSSQKSENLTP 56
492 MKRNNNF---NKDYVDKIFARITYNEKORNNKFSNNNSVYV---QKLKALSYLE 545
QY 57 DOVSOREGIGAEQIVKIDQGYVTSBGHYHYHNGKVPYDA---LFSEEL----- 104
546 D-YSLRKG-----SEKDNHYTLKTGLEADIKLTKTEIKSEKLE 588
QY 105 ---LAKDPYOLKADADYN-EVKGYYI-KYDG---KYVYVLDAAHADVNT----- 149
589 KNFKGLHSANASLEVSQDIYKLOVRYLLIKLIEDLRIELFLKMQLKDSIHVPNIKP 648
QY 150 ---KDEINRQK---EHYKDNKRYNSVA---VASQGRYTTNDGYVF 188
649 QNRPETVYLVLKEDVDKLEPIPKYDMKQDAVLSITQPLVASE---TTEDS--- 702
QY 189 NPADIIDETGNAYIVPHGHHYIPKSDLSASLEAAKAAHLAGKNMOPSOULSTSSVSDN 248
703 ---GSHTHLSQSGELEVETEETVGHHTVITLPPREYVENSISHKNSND- 753
QY 249 NTQSVANGSKSPANK-----SENLSLKEIYD-SFSARYSQSDGLVF 292
754 NSQALFTVYLKLDLFLKSYCHKYILVSNSSMDKLELVNLPDENELKSCDRL-- 811
QY 293 DPAKISRTPNQVAIPHG-----DHYHFIYSK---LSALEEK-IARV--- 332
812 ---DLNINONNIPAMSLDSNANDQHLFEELYOKEMIIYLLKLEENHIKILLPEK 868
QY 333 PISGTGTVST-NAKPREVVSIGLSSNPSSLTSSKELSSASDGYTFNPDVIEETATA 391
869 QITSTSTSPGNTVTWASATHSNQNOSNASS---TWTONG-----VAVSSGPA 918
QY 392 YIVHGHFHRTIPKSNQI-GOPTLPN--NSLATSPSLPIPNPGSHKHEHDGIGFDANR 448

DB 919 VVEESHDPPLTVLISNDLGIIVSLNLTGKTVPNP-LTIST-TEKERYENILKI-MIP 975
QY 449 IIAEDSGFVMSGHNNHFFKKTLEEOIKAAOKLEEWKTS-----HN-----GLDS 497
DB 976 IENDDIKQVKNNSK-----VTGLETOKMALNDEIKKLTQLSFDLYNNKYLKLD 1030
QY 498 LSHEDDYGNNAKEMED---LDKKIEKTAGIKOYGVARE-SIVNKKEMALI----- 547
DB 1031 LNKKELEGDQKQIKLLKLEQESKLSNLPNPNVONFVFNKKKEAIAETWNT 1090
QY 548 -----YFGDHHADPIDBKPVGI--GHSNRYELFRBEVAKKEG 589
DB 1091 LENTKILLNHYKGLVYVNGE---SSPLKLSFVSJOTEDNVALEKFRSLKIDKLAND 1147
QY 590 KYVTGEE-----LTNVNLIKSTFNNOFTLANGORVFSPPPELEKIGIMLV 641
DB 1148 NMLGKRKLSPLSSGHHLLTELEK-EVINKNNT-----GNSPSENNK--VNEAL 1195
QY 642 KL---ITPDGKVLKESGVFGEVGNIANFELDQYLPQGTFRYTI---ASKDYPEV 693
DB 1196 KSEYENLPKAYTVTVTP-----PQDVPSPSLVRVSGSGSKTEETQI 1240
QY 694 SYDGTFTVPLSLAYKNAQOTIFPHAGDTYLVNPOFAVPGTD-----ALRV 743
DB 1241 PMSG-----SLTELOOVLOQNYDEEDSLVPLFSGSENDYLDQVVTGEALST 1294
QY 744 FDEPHNAVLENRYKGEIKLPIPKLNOGTFRAGNKP---VTFMANA----- 789
DB 1295 MDNIIISG--FENEVDIYIK-PL---AGYRSKAKOIEKNITFTNINLNDILNSRLKR 1347
QY 790 ---YLD-----NOSTYIYEVPILEKENQOTRPSILPOFKRNK-----AOE 826
DB 1348 KYLDVLDESIDLMQFKHISSEIIE--DSFRLNSEQNTLLKSYKIKESVENDIDFAOE 1406
QY 827 NSRLDERV---EERTSEKVEKELSETGNSSTSLTEVTVNDPVQAKAFKA--- 877
DB 1407 GISYERVLAKYKDDLESIKVIKE--EKEFPSSPTTPSPAKDEOKSEKFPFLT 1463
QY 878 -----ESYKMKLENTLFNN-----DGTIELYLPSEGV 904
DB 1464 NIETLYNNLVKIDVYLINAKINCONVEKDEAHYKTYLSIDKALDKIDLF----- 1517
QY 905 IKKNADFTGEAPQGNENKPSSENGK-VSTGYVENOPT-----ENKPADSLEAPNE 955
DB 1518 --KNPIDEFAIKLINDTKMDLGRILSTGLVGNPFTIISLISEKFDQMDNISQHC 1575
QY 956 -KPEKPENS 963
DB 1576 VKKQCPQNS 1584

RESULT 9
P3K1.DICDI STANDARD; PRT; 1570 AA.
ID P3K1.DICDI
AC P54673;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PHOSPHATIDYLINOSITOL 3-KINASE 1 (EC 2.7.1.137) (PI3-KINASE)
DE (PTDINS-3-KINASE) (PI3K).
GN PIKA OR PIK1.
OS Dictyostellium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX NCBI_TaxID=44689;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-A3;
RX MEDLINE=96009592; Pubmed=7565716;
RA Zhou K., Takegawa K., Emr S.D., Firtel R.A.;
RT A phosphatidylinositol (PI) kinase gene family in Dictyostellium
RT discoideum: biological roles of putative mammalian p110 and yeast
Vps34p PI 3-kinase homologs during growth and development.*;

```

DR EMBL: U23476; AAA85721.1; -.
DR DctyDb; DD01099; p1kA.
DR InterPro; IPR000341; -.
DR InterPro; IPR000403; -.
DR InterPro; IPR001263; -.
DR InterPro; IPR002420; -.
DR Pfam; PF00792; P1K_C2; 1.
DR Pfam; PF00794; P1K_rbd; 1.
DR Pfam; PF00613; P1K_A; 1.
DR Pfam; PF00454; P13_P14_kinase; 1.
DR PROSITE; PS00915; P13_4_KINASE_1; 1.
DR PROSITE; PS00916; P13_4_KINASE_2; 1.
DR PROSITE; PS50290; P13_4_KINASE_3; 1.
Transferrase; Kinase; Multigene family.
DOMAIN 41 48 POLY-ASN.
FT 59 73 POLY-ASN.
FT 161 168 POLY-ASN.
FT DOMAIN 294 304 POLY-THR.
FT DOMAIN 308 324 POLY-THR.
FT DOMAIN 413 424 POLY-ASN.
FT 503 510 POLY-SER.
FT 570 579 POLY-ASN.
FT DOMAIN 821 828 POLY-ASN.
FT 831 836 POLY-ASN.
FT 1309 1570 P1K/P14K.
SQ SEQUENCE 1570 AA; 178374 MW; 55B678B72B34D783 CRC64;

```

[illegible]

| ID | SC16_YEAST | STANDARD | PRT | 2194_AA | | | |
|------------|---|--|-------|--|--|--|--|
| QY | 469 | -----FKKDLTEBQI----- | ----- | -----RBAOK 483 | | | |
| Db | 459 | KNTNPNQMLSTSQOFLKTLISNEIPIDCKINDINDTDAFSDLSASASSSTPTTSSQS | ----- | 518 | | | |
| QY | 484 | -----HLEEVKTSHGNDLSLSHEDDYPQNAEM----- | ----- | -----KDLCKEIEKIAGIKM--- 528 | | | |
| Db | 519 | LNVQSLRVKAIKTSFNILFLM----- | ----- | -----PQSKKILQVQSGDPIENLKFRRILSDVLFNN 571 | | | |
| QY | 529 | -----QYGV----- | ----- | -----KRESIVNK-----EKNNATY 548 | | | |
| Db | 572 | NSNNNNNNCKYGAADSYLLIDFNDPMNERSVLANKSDYLLDKRAAGLIPKLVEYKSTIL- | ----- | ----- 630 | | | |
| QY | 549 | PHGDHHHDIPDEHKPVGIGSHSHSYELFK----- | ----- | -----PEEGVAKKE----- 587 | | | |
| Db | 631 | -----DSDPSDELSP----- | ----- | -----SEYELIRKLPGTDITMGEEVEYFRFRVTSRLREALP 677 | | | |
| QY | 588 | -----GKRVYTG----- | ----- | -----EELTNVYNLLKNTST-----FNNQNETLA 617 | | | |
| Db | 678 | LKGSIOSTLVRLSPLEPIPVGKKILISLFLPIYQTKRLDELNETADQFNRLFT--- | ----- | ----- 735 | | | |
| QY | 618 | NGQKRVSFSPPELEKKLGINMLVKLITPDQKILEKVSIGFVPEGVGNIANFE-IDQPL | ----- | ----- 676 | | | |
| Db | 736 | -----KNYSKRLPN----- | ----- | -----VSNHFLIKVSSDFJHGPHDRTFESINHHII 778 | | | |
| QY | 677 | PGQTKYTIASKDIPVSYDSTFYVPISLAKYASQITIFPFHAGDTYLRVNDQFAPVPG | ----- | ----- 736 | | | |
| Db | 779 | QSTKQPLTLQRPREDLPQ----- | ----- | -----KPRFDYPP 809 | | | |
| QY | 737 | TDALVRVDEDFHGNAYLENN----- | ----- | -----YVGSIKLPIPKLNGCTR 775 | | | |
| Db | 810 | L-----IIDHCSNAINCNNNNNTNNTNNINFDNDQITHSIREIKKPFKRVYWGSR | ----- | ----- 864 | | | |
| QY | 776 | TAGNKIPYTFMANYVLNDQSTIYVEVILEK----- | ----- | -----ENQCDKPSILPQFK----- 820 | | | |
| Db | 865 | -----IPLSCIKDI----- | ----- | -----DSSVIVSISLYHGICFSKATQPIIPPFPAFLKTLSDVDC 916 | | | |
| QY | 821 | -----RKKAQNSKLDERVEEPTSEVVEKEKISLSTGNSTNSLSEPTVDVPOEKYAK | ----- | ----- 875 | | | |
| Db | 917 | EMLVFTNIDYSNLPPDARL----- | ----- | -----SISVYSANETVDVDE----- 951 | | | |
| QY | 876 | FAESYGMKLENVLFNMDGTTELYLPSEGVYKKNMADFTGEAPOGNEKN----- | ----- | -----PSENGKYST 933 | | | |
| Db | 952 | -----INLDEATKKLLEPIG-WINWMITPFKYQLRGAYVELSLMPDSN-PL | ----- | ----- 997 | | | |
| QY | 934 | GTVENOPTENKPADSLPE----- | ----- | -----APNEKPYKPNSTNDGML-NPEGVNGSDPMLPALKE 987 | | | |
| Db | 998 | GTCSNPNSSQSVGLTEFEPEFNLPLYLPFKRTKFTFSVSVIYQPPNTNINSNMBE-FFEQ | ----- | ----- 1056 | | | |
| QY | 988 | APAVDPVOE-KLEFFTASLYGLGDSYTF 1014 | ----- | ----- | | | |
| Db | 1057 | ITALDPLSLDKOEKYNQLMTLRHSILF 1084 | ----- | ----- | | | |
| RESULT 10 | | | | | | | |
| SC16_YEAST | | | | | | | |
| ID | SC16_YEAST | STANDARD | PRT | 2194_AA | | | |
| AC | P48415 | | | | | | |
| DT | 01-FEB-1996 (Rel. 33, Created) | | | | | | |
| DT | 01-FEB-1996 (Rel. 33, Last sequence update) | | | | | | |
| DT | 15-DEC-1998 (Rel. 37, Last annotation update) | | | | | | |
| DE | MULTIDOMAIN VESICLE COAT PROTEIN. | | | | | | |
| GN | SEC16 OR YPL085W. | | | | | | |
| OS | Saccharomyces cerevisiae (Baker's yeast). | | | | | | |
| OC | Eukaryota; Fungi; Ascomycota; Saccharomycotini; Saccharomycetes; | | | | | | |
| OC | Saccharomycetales; Saccharomycetaceae; Saccharomyces. | | | | | | |
| OX | NCBI_TaxID:4932; | | | | | | |
| RN | | | | | | | |
| RP | | | | | | | |
| RA | SEQUENCE FROM N.A. | | | | | | |
| RA | MEDLINE:96017704; PubMed:7593161; | | | | | | |
| RA | Bependshade P., Glimeo R.E., Holzmacher E., Teng P., Kaiser C.A.; | | | | | | |
| RT | "Teat SEC16 gene encodes a multidomain vesicle coat protein that | | | | | | |
| RT | interacts with Sec23p." | | | | | | |
| RL | J. Cell Biol. 131:311-324(1995). | | | | | | |

CC -1- FUNCTION: INVOLVED IN THE BUDDING OF TRANSPORT VESICLE FROM THE
 CC ENOPLASMIC RETICULUM. THE C-TERMINAL INTERACTS WITH SEC23 AND
 CC WITH THE CYTOSOLIC DOMAIN OF SED4. COULD THEREFORE BE A
 CC CONSTITUENT OF COP1 VESICLE COAT. N-TERMINAL OVEREXPRESSION
 CC CAUSES A LETHAL SECRETION DEFECT.
 CC -1- SUBCELLULAR LOCATION: ON THE ENOPLASMIC RETICULUM AND ON VESICLES
 CC WHICH BUD FROM IT.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).

DR EMBL: U23819; AAC49088.1;
 DR SGD: S0006006; SPC16;
 KM Transport; Protein transport; Golgi stack; Endoplasmic reticulum.
 FT MUTAGEN 1058 L->S; IN SEC16-4; TS ACCUMULATION OF ER
 FT MUTAGEN 1083 1083 MEMBRANES.
 FT L->P; IN SEC16-3; TS ACCUMULATION OF ER
 FT MEMBRANES.
 FT MUTAGEN 1088 1088 L->P; IN SEC16-2; TS ACCUMULATION OF ER
 FT MEMBRANES.
 FT MUTAGEN 1230 1230 M->R; IN SEC16-1; TS ACCUMULATION OF ER
 FT MEMBRANES.
 SO SEQUENCE 2194 AA; 241613 MW; BB1E02D2AD4683E3 CRC64;

Query Match 3.4%; Score 185; DB 1; Length 2194;
 Best Local Similarity 21.6%; Pred. No. 0.61;
 Matches 223; Conservative 124; Mismatches 418; Indels 266; Gaps 54;

QY 92 GAVYDALESEELMDPYOLKADIVNEVKG-----YIKVDGK----- 133
 DB 1282 GNP-----LSNEP--VADSDVIFESTGNMTFESILMDEIYEIYFSDPRFKGF 1330
 QY 134 -----YYVYLKDAHAADNVTK--DEINQO-KOEHVNDKENV-----SNVAVARSO 177
 DB 1331 SSILPQKIYHASILQEGNLSLQTKYTDYLSSEVSRKLPKDKILTIMTRELSEVASRLSE 1390
 DB 178 GRITTDGVVNP-----DIIEDGNAYIVPHG-----HYHYIKRSDI 217
 DB 1391 ----SWTGMIAKRLSSVWGQIDKRSFKYI--GGDDIALNKKNDKKVDFDGTGSSA 1443
 QY 218 SASELAAAHLAGKNMOPSLSYSTASDNMTOSVAKST--SKPANSENLQSLKEL 275
 DB 1444 NSSTVDLQTFPPFOAQVTSQSYVDTALLNMAHNPVSHVLSKTSNVS-----KGL 1496
 QY 276 YDPSAQRSESDGLVFDPAKIISKTPNGVAIPR-----GDHYHETIPYSKLSALEE 326
 DB 1497 VEANLPYTHRIDSLQSPQR-IHNTQFAALEPOMASLRKRYRDQHTNEKALKSOQILKE 1555
 QY 327 KIARMPISGSGTSTNAKPNREVSSLSGSLSNPSSILTSKLSASDSOYIINPDIYE 386
 DB 1556 KSTAYTPQFQONHSPKESNSNVPSTLEADFPAP-----RLGTVPSNVTSSP-DLVR 1607
 QY 387 ETATAYIVRGDHFHYIPKSNQIGOPTLPNNS-----LATPS-PSLIPNPTS--HEKHEE 439
 DB 1608 RES---TISGSEFLPPRK--IGVTPKANSOGSLMTSSVRLPIDVPVQVHERGYN 1661
 QY 440 DGYGFDAARIIADES--GFVMSGHGHNHYFFKKDLTEQIKAAQKHLVEYKTSNGLDS 497
 DB 1662 DFGKHSQKSMPEDESHSHDNSADONTLKDSADVDDEFMDIEGPGFNDVK--NLTPM 1718
 QY 498 LSSHEDYPRNAEKMDLKIIEKTAGIKQGVGRRESIVNKKKALIIYPRGDHHD 557
 DB 1719 EPNHQPSTVNPLO-----TISDDIPII-QTNVEVRGTDASQWENSL--PSIENERS- 1768
 QY 558 PIDEKRVVIGHSNHYELKPEEGVAKKEGNKYTGEEELTNVNLKSTFNNOQTLLA 617
 DB 1769 --SEEPENISKSNSS--ATLPSTGGISLE-NRPLTODE--NSISEVQOSTY-----LP 1815

QY 618 NGOKRVSEFPELEKLGINKLVLTIPDGVNLEKVGATFGEVGNIA----- 667
 DB 1816 AG-----SISWEAKPISQVDPVPRNVNKKSKLVEQHNAPPRKSTADT 1859
 QY 668 --NEFLDPIYLGQFKTKTASKDQPEVSYDQTFVPPVSLAKMASO-----TIFYPP- 718
 DB 1860 KKNY---SPYVP-----QSTVMSADDE---STIIKTSATYARHQAHSNPQTFPLV 1908
 QY 719 -HAGDTYLRVNOFAVPGTDALVAVDFEFGNATLVNRYKGEIKLPKLN----- 770
 DB 1909 NQANET-----ASFELSESTSOA-----QSNQNVASENRFS-----PIKALVEYKDTF 1952
 QY 771 QGTTTAGNR-----IPVYFMANAYL-----DNOSTYVEVPILEKENDQDPSIL 816
 DB 1953 QPTIKASINQVRAKPLESDAKTNDYDESDDDNNSDTDAKKNRKKVNNMKETK 2012
 QY 817 PQFRNRKAQENSKLDEK-----VEEPTSEKVEKEKISEFGNSISNTLEVPYVDV 869
 DB 2013 P-----SNKDIDKSGNFGWGLKQDYGDKKVKAKIGHKNTLYYDEKLKRWVNRDAT 2064
 QY 870 QEKVAKFASVGMKLENTLVNMD-----GTIELYLP-----SGEVIRKNADFTGE-- 915
 DB 2065 EEKOKIIESSAPPPPIYKRRDGGPKTKRNSGPIINSLPVAHVSIVYNN--PTGEPD 2122
 QY 916 -----APQG-NGENKPSNGKVS--TGTVENQPTENKPADSLPEAPNEKP--VKPENST 964
 DB 2123 PIKTSPPSPGPPNPNNSPSPSPISRLSGVNLASKANGDLDLSLAGGPRPASTRAKKKT 2182
 QY 965 DNGMLNPEGNV 975
 DB 2183 ARGVAVVMDNI 2193

RESULT 11
 SP2 YEAST
 ID SP2201; STANDARD; PRT; 1466 AA.
 AC P23201;
 DT 01-NOV-1991 (Rel. 20; Created)
 DT 01-NOV-1991 (Rel. 20; Last sequence update)
 DT 30-MAY-2000 (Rel. 39; Last annotation update)
 DE SP22 PROTEIN
 GN SP22 OR PEA1 OR YIL021W OR YIL209.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_Taxid=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=8288C;
 RX MEDLINE=91009481; PubMed=2211820;
 RA Gehring S., Snyder M.;
 RT The SP22 gene of *Saccharomyces cerevisiae* is important for
 RT pheromone-induced morphogenesis and efficient mating.
 RL J. Cell Biol. 111:1451-1464(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=8288C;
 RA Fumelle B., Goffeau A.;
 RL Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: INVOLVED IN PHEROMONE-INDUCED MORPHOGENESIS AND
 CC EFFICIENT MATING. PERHAPS AS A CYTOSKELETAL PROTEIN.
 CC -1- SUBCELLULAR LOCATION: LOCALIZES A SHARP PATCH AT THE SHMO TIP
 CC (MATING PROJECTION) WHICH IS THE SITE OF POLARIZED CELL GROWTH.
 CC -1- SIMILARITY: TO PROTEINS THAT FORM COILED-COIL STRUCTURES.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).

```

CC      EMBL, X53731; CA37763.1;
DR      EMBL, X91560; CA366170.1;
DR      EMBL, Z73126; CA367469.1;
DR      SCS: A36426; A36426;
DR      SCS: S0003944; SPA2.
KW      Cell shape; Repeat; Coiled coil.
FT      DOMAIN 286 388 COILED COIL (POTENTIAL).
FT      REPEAT 818 1087 25 x 9 AA APPROXIMATE TANDEM REPEATS.
FT      REPEAT 825 834 1.
FT      REPEAT 826 834 2.
FT      REPEAT 835 843 3.
FT      REPEAT 860 868 4.
FT      REPEAT 875 883 5.
FT      REPEAT 884 892 6.
FT      REPEAT 893 901 7.
FT      REPEAT 902 910 8.
FT      REPEAT 911 919 9.
FT      REPEAT 920 928 10.
FT      REPEAT 929 937 11.
FT      REPEAT 938 946 12.
FT      REPEAT 947 953 13.
FT      REPEAT 954 961 14.
FT      REPEAT 962 970 15.
FT      REPEAT 971 979 16.
FT      REPEAT 980 988 17.
FT      REPEAT 989 997 18.
FT      REPEAT 998 1006 19.
FT      REPEAT 1007 1015 20.
FT      REPEAT 1036 1044 21.
FT      REPEAT 1045 1053 22.
FT      REPEAT 1054 1062 23.
FT      REPEAT 1072 1080 24.
FT      REPEAT 1081 1087 25.
FT      DOMAIN 1169 1189 COILED COIL (POTENTIAL).
FT      DOMAIN 1275 1302 COILED COIL (POTENTIAL).
SQ      SEQUENCE 1466 AA; 163142 MW; 2EBB61513282C89 CRC64;

```

Query Match 3.4%; Score 184.5; DB 1; Length 1466;
 Best local similarity 18.2%; Pred. No. 0.37;
 Matches 204; Conservative 163; Mismatches 386; Indels 369; Gaps 49;

```

QY      24 ALNQHRSQENKDNRYVYDGSOSQSKSENLT---TPDQVSQK-----GIOAE 68
DB      316 SLTSEKEKEKEE-----KEEKEKENKINNTYIDSPKRELLSLNSQGEISTENE 367
QY      69 QIVIKITD---QGYTSHGD-----HYHYNGKVPYDALPSEELMKDPYOLK- 114
DB      368 NLKOKISEELHOKKNDNDKLTGDFISKYSADGLIPAOYILNANLLIIOFTTRLSA 427
QY      115 ---DADIYNEVKGIIIVDQKRYVYLKD-AAHADNATKDEINROKOEHVKNERNVS 169
DB      428 VIGGISTASHOIGELRPIISQSLNLSQLLSLADLQYKQVILK-----A 476
QY      170 NVAVASOGKRYTNDGYFNPADIIEDGNATYPRGHGHHYIPKSDSA--SELAAKA 227
DB      477 SLSHAITSIRYS---YTGPFV-----LIPKITVOANVSEVCFAMC 513
QY      228 HLAGKMMQPSQISYSTSDNNTQSVANGSTKPKANKSENLQSLKELYPDSPAORYSES 287
DB      514 NL-----IDSAKIKSDSNGES---TTSNEGNR-----QVLE--YSSPTAL- 548
QY      288 DGLVDPKAKIISRTNPGVAIPRGDHYHTI---PYKISALAEKATARAVPI----- 334
DB      549 ---TPPTPTPTPTSGINMKKG---PINPKPASFLNDEEESPPVKPKITOKAIN 599
QY      335 ---SGTGSYTVNAKEN-----EYVSSIGSLSS---NPSLITTSKELSSADGYI 378
DB      600 PIIRSSSNGVPTTSKRKSGGLSLIMDSSIAKNSKRENDKTVSKIKATVTSASNSAS 659
QY      379 FNPADIYEETKATYVRHGDFHTIPKSNQIGOPTLPNNSLATPSPSLINPGTSHERKE 438

```

```

DB      660 SNISEIRKLT-----LPPQAKIGTVIPSENOV-----PIKIENTE 696
QY      439 EDCIGDPANRIIIEDESGYVMSGHNNHFFKKDLTEE-QIKAAQHLEEVKTSNGLDS 497
DB      697 ED-----NK-----RSDIINEISVKTSPSSIAKTKOFENSSSEK 729
QY      498 LSSHEDYPPENAKEMDQDKIEKTAGIMKQGVARESVVAKENNAIIPRGDHHND 557
DB      730 KSSPKEN--PIAKEEDSKRISNKTITSNDVSTDDSSDGN-----ENDDAD 776
QY      558 PIDEHNPVIGISHSNVELFKPEEGVAKREKNRYTGEEL---TNAVNLKSTFNQNF 614
DB      777 DDDDFYML-----KQTMKRGSKIEKNNSKLPANVELDLIESPSVNI 823
QY      615 TLANGORVSFPPELEKLGIMLVKLTTPD-----GVLEKYSVYVGBGNGN 665
DB      824 ESSESEKETS---EMSEMPSSSLPKRLVEDEVESEPEKASVSEYKKNFQELGN 880
QY      666 IANFELDO-----PYLPQGTFFYTLASK-----688
DB      881 VESPDNTQVKSIGMTGKAVGPESDSRVSPGNTGQIKSLNNAKGVVPEADSRVSPGM 940
QY      689 --DYPEVSYDGEFVPTSL-----AYKASQTIFFYFHAGDYLLRYVQFAVPKGTDALY 741
DB      941 KEOIKSIGMTGKTQESIKSPEAKRLAS-----SGEVDKISPRV--RESESL 990
QY      742 RYFDEHGNAYLENNKTVGEIKLPKLNQGTTRFAGNFIPTVMANAYLNDQSYIYEV 801
DB      991 AV-----GNTIPSNM---TVKMESEPL--KGVTVSPQDRIIDLSSEPIEN----- 1032
QY      802 PLEKENQDPSILPQFRNKAQENSKLDEVEEKEKSEKIEKLESTGNS-----T 855
DB      1033 ---VDPKVLAKIYFPFAVNTGSPKSEVETPSATLNSGLPEPNSQIYSPELA 1084
QY      856 SNSTL-----EYVTVDPVQE-----KVAKFAESYGMK 883
DB      1085 KNSPLAPIKKNVELRETNNPHETITSVEPTKNDANTSWRODLNRTIKREDEDDDFRY 1144
QY      884 LENV---LFPNDGTIELY-LPSGEYIKNMADFTGEAPQENGENKPEENGSTGTVE- 937
DB      1145 NNINIOITGAYTGTGIDYKIPVDKRAKSEAEVHTSEEDJDESN---VNGRRDAQCHI 1201
QY      938 ---NOPENKPADSLPEAD---NEKPYKRENTDQNLN 970
DB      1202 TERKAHVPTENSOVKNTSHPPLNSKPYOVENSESNGIN 1243

RESULT 12
BUDJ_YEAST STANDARD; PRT; 1636 AA.
AC P23558; P23557; P87007;
DT 01-NOV-1992 (Rel. 22, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE BUD SITE SELECTION PROTEIN BUD3.
GN BUD3 OR YCL014W OR YCL14W/YCL13W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Chant J.S.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Oliver S.G.; Anwar R.; Brown A.; Gent M.E.; Indge K.J.; James C.M.;
RA Statera L.L.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Gromadka R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.

```


CC - FUNCTION: CO-ASSEMBLES WITH BUD4 AT BUD SITES. BUD4 AND BUD3 MAY
 CC COOPERATE TO RECOGNIZE A SPATIAL LANDMARK (THE NECK FILAMENTS)
 CC DURING MITOSIS AND THEY SUBSEQUENTLY BECOME A LANDMARK FOR
 CC ESTABLISHING THE AXIAL BUDDING PATTERN IN G1
 CC - CAUTION: REF. 3 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 1338
 CC ONWARD AND IS SHORTER (1367 AA) DUE TO A FRAMESHIFT
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: 017580; AAA66315.1; ALT_FRAME.
 DR EMBL: X59720; CAA42346.1; ALT_FRAME.
 DR PIR: S19341; S19341.
 DR PIR: S19340; S19340.
 DR SGD: S0000520; BUD3.
 DR CONFLICT 285 Q -> P (IN REF. 3).
 FT CONFLICT 407 R -> N (IN REF. 3).
 FT SEQUENCE 1636 AA; 184717 MW; 9E4E6BA5C3A3F69 CRC64;

Query Match 3.3%; Score 179.5; DB 1; Length 1636;
 Best Local Similarity 18.7%; Pred. No. 0.72; Indels 347; Gaps 47;
 Matches 186; Conservative 135; Mismatches 327;

91 NGVPPDADFSEELKKDDPKYOLKADIVNEVGGYIIKVDGKYYVYLKDAHADVNRK 150
 722 NSSMADLLCVNENLKNHEHOLEV----- 747
 151 DEINROKOEHVKNENKVNVAVA-----RSQRYTNDGYVFNPAIDIEPTGNA 200
 748 -----KHPTDEHRAVNSKLSGASPDAMHEKRSYGITTRSTSLDKD----- 793
 201 YIVPHGGHYHYPKSDLSASELAARAHLAGKMKOPSQSLSTSTASDNTQSAKSTS- 259
 794 -SPSGDMSVTKET-----KEILPYKPKSKSKPKREIÖKTKTKTNASK 835
 260 -----KPAKSENLSLKELDSPBAORYSDDGLVPPAIIISTPFGVALPHGDHY 313
 836 AEHIEKKPKPKGFGVGLNVGSKSKSPSVÖRV--PKISÖRHPS----- 883
 314 HPIPKSLALEKIAWPIISGSGVSTNAKPNVYSSLSGSSNPSSILTSKELSA 373
 884 ---PVKPKPTSEK-----SSPKRAVSSPKTKKSTSEST-KESQTA 922
 374 -----SDGIYFNPKDIVEETATAYIVRHGDHFHYIPKSNQIQÖPILPNNSLATPS 423
 923 KSLRAVEFKSDDLIGRPD-----VGNCAH-----POEWTRISSVYRÖKRYSYN 968
 424 PSLDIPNGTSHKEEDGEGFDANKRIIAED-SEFVASHGDHNIYFKKDL-TEÖI--K 479
 969 PSÖPTEVTSNEKNVEP---KADÖSTKÖDNISNFADEVASASSYPRKILAEVTDÖITIGK 1024
 480 AAKR-----HLEEVTSHGIDLSHEDÖYPGNAEKMDLD-----KRI 519
 1025 ATNSSVHGNKELPDLEEVTA-NRVSTTSAGDÖRIDTÖSEFLAAYDENISDDÖHRÖN 1083
 520 EEKIG-----IMKÖYGVRESIYVKNENKNAIYIPGHGHHADPIDÖKRVGIGSHS 572
 1084 ESRVNDOLFÖDÖFIPKHNKNÖENI--NSSN--LFEPEG-----KVQÖEK--CVSNENT 1131
 573 NYELKPEEGVAKKEGNTVYGEELNVVNLKSTFNO-----NFLANÖK 621
 1132 NISLKTNDASTLÖKLSPOASKVLTENSNEIKDT--NNEKDAKIDKLGÖYDKÖETAK 1189
 622 RVSEFPE-----LEKKIGIMLVKLTIPDGKYLEVSGKRVGEVGNANFELDÖPY 675
 1190 ETT---KKNFVÖGTEKKEIFPTIPRLAPASKI-----NÖRÖRSY 1229

676 IPGÖFKTIASKYPERVSYÖDTFVPSIATKASQTIFFHAGDYLAVNPOFAVRK 735
 1230 I--ELFÖGKRVILDKHADYIMK-----RLASÖVL-----SEGLKVN----- 1265
 736 GUDALVAVDEHNGVAYLENKIVGEI-----KLPPKLNÖTTRAGKRIPTVFANRY 790
 1266 -TEEDAAIINKÖDÖKAKERMÖISEVYIEMÖÖP-----TYIPKHA 1308
 791 LNOÖSTYIYEVILEKENÖTDPKPSILPÖKRNKADÖNSKIDKEVPEPKSEYENKELSE 850
 1309 LDDSG-----IEKSD-----DFFELIEELKEELGSKTGNDVON 1344
 851 TÖNSYNSSTLEVPY-----VDPÖEKVAKFASISGMKLEN-----VLNMD----- 892
 1345 NNPNSIPIKIEKPPAKVYRTSPVRIIGTFEDT--RKTENGSPSDISFTYDHNDEPD 1402
 893 -GTIELTSPGVIKKNADFGEAÖNGENKPSNGKVSCTGVNÖPTENPADSLPE 951
 1403 KRLMEKTFPSÖDEIPDDR-YT-----PAE-----EPTAEFPVELÖPN 1439
 952 APNEKFPKPEK--STÖNGMILNPEGVNSDP--MLD 982
 1440 TRPSIVTTSNKNSTÖDKL--SSGNDÖKPTÖELD 1472

RESULT 13
 IGK2_HAEIN STANDARD; PRT; 1702 AA.
 ID IGK2_HAEIN
 AC P4384;
 DT 01-NOV-1995 (rel. 32, Created)
 DT 01-NOV-1995 (rel. 32, Last sequence update)
 DT 01-NOV-1995 (rel. 32, Last annotation update)
 DE IMMÖGLOBULIN A1 PROTEASE PRECURSOR (EC 3.4.21.72) (IGK1 PROTEASE).
 GN IGA.
 OS Haemophilus influenzae.
 OS Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OC NCBI_Taxid=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HK715 / SEROTYPE B;
 RA MEDLINE=92234949; PubMed=1373717;
 RA Poulsen K., Reinholdt J., Killian M.;
 RT "A comparative genetic study of serologically distinct Haemophilus
 influenzae type 1 immunoglobulin A1 proteases.";
 RL J. Bacteriol. 174:2913-2921(1992).
 CC - FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMÖGLOBULIN A
 CC - PRODUCING INTACT FC AND FAB FRAGMENTS.
 CC - CATALYTIC ACTIVITY: CLEAVAGE OF IMMÖGLOBULIN A MOLECULES AT
 CC CERTAIN PRO-1-XAA BONDS IN THE HINGE REGION. NO SMALL MOLECULE
 CC SUBSTRATES ARE KNOWN.
 CC - SUBCELLULAR LOCATION: SECRETED.
 CC - DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
 CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
 CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
 CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
 CC SIMILARITY). BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: M87489; AAA24966.1; -
 DR MEROPS; S06.001;
 DR InterPro; IPR000710;
 DR PRINTS; PR00921; IASERPTASE.
 KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
 FT SIGNAL 1 25 POTENTIAL.

FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
 FT PROPEP 1015 1702 HELPER PEPTIDE (POTENTIAL).
 FT ACPT_SITE 288 PROBABLE.
 FT DOMAIN 1109 1124 2 X 8 AA TANDEM REPEATS OF A-K-V-E-K-E-E-
 FT REPEAT 1109 1116 K.
 FT REPEAT 1117 1124 K.
 SQ SEQUENCE 1702 AA; 186539 MW; 860F70D2667807A6 CRC64;

Query Match 3.3%; Score 177; DB 1; Length 1702;
 Best Local Similarity 19.6%; Pred. No. 1;
 Matches 223; Conservative 144; Mismatches 430; Indels 338; Gaps 53;

QY 84 GDHYYHNGRPV-----DALF-----SEE-----LAKDP--NQ-LKAD-----IV 119
 DB 236 GDAYTGIAGTPYKVNHEHNGLLGFNGSKREHSDPKGILLSDLLTNAYVAGDSGSPFTV 235
 QY 120 NEVKGCIIVDQKYYVYLKDAHADNVTKDEINRQKHVYDNEKVNVAVASQGR 179
 DB 296 DREKGMWLL--GSY-----DFWAGYKRSQWEMNITYKEPAK-----TVLDKDTAGS 341
 QY 180 YT-TNDGYVNPADIIEDTGNAYIVPHGHYIIPKSDLSASFLAAKHLAKNNQPSQ 238
 DB 342 LIGSNQYKNNP-----IGKTSVINGSESLNVDLP--SSODTSKNNHRSV---T 390
 DB 239 LSTSTASDNN-----TQSAKSTSKPAKSENLSLILKELYDPSAQ 282
 DB 391 LRSGSTLTNNNIDQAGGLFEFGDEYVGTSDSTTWKNGSVADGKTYVWVHPKSD 450
 QY 283 RYSE-SDGLVFPDAKISTRTPNVAIPRODHFHFIPYSLKSALEKRIARVPISGSTV 341
 DB 451 RLAKICKGTLI-----VEGGENKSLKVDGTVILQADANKKVAFAQGVIGRSTV 506
 QY 342 STN-----AKPNEYVSSL--GSLSSNPSSLTSKELSSASDGYIFNKDIYEATAYIR 395
 DB 507 VANDQVNPNSIYFRCGRGRDANNNL--TEHINRIDG-----ARLVN 551
 QY 396 HGDHFIYIKSNQIGPPLPNSLAT--PSPSLPIPGTSHE-----KHEEDG----- 441
 DB 552 H-----NTSKTSVITIGESLIDPTIIPYNDADADDENPFAFRIRKGGOLYLWL 603
 QY 442 --YGPANNTIADDESGFVMSGDHN--HFFPKDLTEQIKAAQKHLVEKTSHNGDLSL 498
 DB 604 ENYTYALAKGASTSELPKNSGSENNMLYMGKTSDEAKRMVHHINERR--NGFNQY 661
 QY 499 SSHEQ-----DYPGNAKEMKDL-----DKREE-----KIKGI 526
 DB 663 PEEEGKNNGNLNTFKKSGENRLLGGTVNLNGDLVKEGTLFLSGRPFHARDIAGI 721
 QY 527 MK-----QYGVKRESIV-----NKEKNALIIYPHGDHHD--PIDEHKPV 565
 DB 722 SSTKDKQFAENNEVYEDDMINRFKATNINVTNNATILYSRNVANITSNTIADNAKY 781
 QY 566 GIGHSNVELEKPEEGVAKKGNV-----YTGEELNVNVLNLSNFPNNQFTLANGQ 620
 DB 782 HIGY-----KAGDVCYRSDYTG--YVCTDIDKLSDKLNSFNATNATVGN 824
 QY 621 KRVSEFPPELEKRIGINMLVLLTPDG-----KYLEKVSIGVFGGVGNINAFELDQYL 676
 DB 825 VMLS-----GNANFVLGKANLFCITISGTNSOVRLENSHMHLLTGSNNQNLIDGHIHL 880
 QY 677 PEGQFKYTIASKD---YEVSYDGTFTVPTSLAYKMASQTLFFPHAGDTYLKVPQPAV 733
 DB 881 NAQNDANVTYNTLVNLSGNSGFYLLDLSNKGQKVVYVTSANGNFLQVADKIGE 940
 QY 734 PKGDALVRVDEPHGNAVLENNKTVGEGIKPIPLRNLQGTTRTNAKTPVTFMANAYLDN 793
 DB 941 P--TKNETLTLD--SNA--TRNNINVSIV-----GNTVDLQAMKYLKLVN 980
 QY 794 OSTIIVEPILIKENQ-----DKPSI-----LPQF 819
 DB 981 NGRDVLNFEVBEKKNQYVDITNTTPNNIADVPSVSNNEELARVETPVPPAPAPPSE 1040

QY 820 KRNAQNSKIDK-----VEEPKTEK--VEKEKLSNGNSNS 858
 DB 1041 TTEVAENSOESTYKKNEDATETTAONGEVAEKPVSANTQNEVAQSSSETEET 1100
 QY 859 TLEEVPTVDQV--EKVAKFALSTGKLENTLFFNNDGTEILYLPBGVYKKNMADFGEAP 917
 DB 1101 QTEIKETAVKEKEKAKYKEKAKYK-----DEIDQAP 1136
 QY 918 QGNGENP-----SENGVSTGVENOP-----TENKPADSL--PE 951
 DB 1137 QMASSESPKQAKPAKPEVSTDTKVEETQVOAQOTOSTTVAALATSPNSKPALETPSE 1196
 QY 952 APKEKPKP-----ENSDNGMLNPEGANGSPMDPALKEAPV---DPQEKLE 999
 DB 1197 KINAEPTVPVSKNQTENTDQPTEREKTAKEVTEKTOEPPQVQASQAPROOSE 1251

RESULT 14
 HMW1 MYGE
 ID HMW1 MYGE STANDARD; PRT; 1139 AA.
 AC Q49413; Q49365;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE CYTADHERENCE HIGH MOLECULAR WEIGHT PROTEIN 1 (CYTADHERENCE ACCESSORY
 DE PROTEIN 1).
 GN HMW1 OR MG312.
 OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Molluscites;
 OC Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2097;
 RN [1].
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 33530 / G-37;
 RX MEDLINE=96026346; PubMed=7569993;
 RA Frazer C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
 RA Nguyen D.T., Uterback T.R., Sauder D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.-F., Dougherty B.A., Bolt K.F., Hu P.-C., Lucier T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,
 RA "The minimal gene complement of Mycoplasma genitalium.";
 RL Science 270:397-403(1995).
 RV [2].
 RP SEQUENCE OF 721-847 FROM N.A.
 RC STRAIN-ATCC 33530 / G-37;
 RX MEDLINE=94075230; PubMed=8253680;
 RA Peterson S.N., Hu P.-C., Bolt K.F., Hutchison C.A. III,
 RA "A survey of the Mycoplasma genitalium genome by using random
 sequencing.";
 RL J. Bacteriol. 175:7918-7930(1993).
 RL J. Bacteriol. 175:7918-7930(1993).
 CC -1- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
 CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
 CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
 CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHERIN PROTEINS
 CC IN THE MYCOPLASMA MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: LOCALIZES SPECIFICALLY TO THE ATTACHMENT
 CC MEMBRANE (BY SIMILARITY).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC
 CC EMBL: U39712; AAC71534.1;
 CC EMBL: U02261; AAD12527.1;
 CC TIGR: MG312;
 CC Cytadherence; Structural protein.

SEQUENCE 1139 AA; 130531 MW; 0011D3288C3DD856 CRC64;

Query Match 3.3%; Score 176.5; DB 1; Length 1139;

Best Local Similarity 18.6%; Pred. No. 0.6; Indels 247; Gaps 39;

Matches 160; Conservative 129; Mismatches 322; Indels 247; Gaps 39;

265 SENTASLRELKELDPSAQRYSSED-----GLVEDPAKIIISRPN-GVAIPHGDIHYFIY 318
 DB 2 ANKQSVFE-----KNYTOEPENINGDLYDKSTVEDPNIKAADADANGYIAF 54
 319 SKISAL-----EETIARVPISGTGSTVSTNAPNEVSSLSLSPSSLTSTKEL 370
 55 NKESTGYIYPYGDTEYDISQLFEDENGPNPFDEKOEENDYLKTVG---NPDGSDYENG 110
 371 SSSADGTYIPNDIYETATAYVRHGDHFIYIPKSNQIGOPTLPNNLSATPSPSLP 430
 DB 111 EWWSYTFEDDOWI-----STKESQPTDENTGFDSDLP-----P 144
 431 GTSHEKHEEDYGFDPANRI-----IAEDSGF-----VMSHGDHNYFP 469
 DB 145 EVKQPSVEEDNCFD-NOLPREVKQPSVEDNNGFNDLPREVKQPSVVDQSSSDYFA 203
 470 KKDITEBOIAPAAOKHLEEVTSNGIDSL-----SHEODYFGNAKEMK 513
 DB 204 KOP-TDENGFNDLPREVQPSVVDQSSDDHFAKQPESTYDSFDLPDQPTLDQ 262
 514 DLKTEKTEKJAGIMKOYQVARESVIVNKENALIIYHGDHNDPIDEHNPVIGSHSN 573
 DB 263 SLDDHV-----QYNF-----DHR-----EELKPV-----AERON 287
 574 YELFKPEGVAKKKNVYVGEELTVNVLKSTFNQNTLANQKRVSEFPPELEK 633
 DB 288 YOV-----GDOVOAN-LDNNEIOPAEKKTTPDESQAQVVD-----SYGLPIDTQ 336
 634 KLGIMNLVKLIT-PDGKVLKESGVKGVGEGVGNANFELDQPLPGGTFYTTASKDYE 692
 DB 337 QDQTFSSFEQPTVQEQFOVNSEV-----NDQFPEITKEVYLE 377
 693 VSDGTCTVPTSLAYKASQITIFPFAAGPTILRVNQAFAVPGTDLAVVF-----744
 DB 378 SSEFKQDVETS--DINSSENLSENKDA-----INDSLSEFIOLNSENSE 423
 745 ---DEFHGNALYLE--NNYKGEIKLPIPKLNOGTTTRAGNKIPYFNA---NAVILNOS 795
 DB 424 TASDVHYEKSESEIHRYKGS-----DLSQNSNNSLSEPKFSEAPDAHFESQS 477
 796 TYI--VEVPILEKEN---QTDKPSILFOFKRKAQENSKLDER-----VEEPT 839
 DB 478 EPVQVQYDIYQNELKPTLDQPSDDYFAKQPTDENTGFDNDLPREVQPSVVDQSS 537
 840 SEKVEKELSTGNSNSTLEEVPTVD-----VQ-----EKVAKFAESYGMKL 884
 DB 538 DHRAPQPESTDSYSDSDLPQ-PTLDQPSLDHVOYNFDHHEELPVABEONNYGVGF 596
 885 ENVLFNDGCTETLPSGEVYKKNMADFTGEAPQNGENKPSSENGKSTGTVENQPTENK 944
 DB 597 DOVQANLDNNEIQ-PTAE--KEYTTPDESQAQ-----VYDSYQPLID-DOQOQT 645
 945 PADSLPEAPNPKPYKPNSTNDGNLNEGANGSDPMLDPALEAPAVDPOVEKLEKTFTAS 1004
 DB 646 FSSSEPOPTVEGPDQVNSEVNDQFKE--ITKEPVLESSEPNKODVVE-----TSN 694
 1005 YGLGDSYTFNMQCTIEL 1022
 DB 695 YTNMLQKEDIGSDNKITI 712
 RESULT 15
 IGA_NEIGO STANDARD; PRT: 1532 AA.
 AC P09790;
 DT 01-MAR-1989 (Rel. 10, Created)

01-MAR-1989 (Rel. 10, last sequence update)
 DT 01-OCT-2000 (Rel. 40, last annotation update)
 DE IGA-SPECIFIC SERINE ENDOPEPTIDASE PRECURSOR (EC 3.4.21.72) (IGA
 GN PROTEASE)
 OS Neisseria gonorrhoeae.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 NCBI_TaxID=485;
 RX MEDLINE-90154052; PubMed-2105953;
 RA Bachovchin W.W., Plant A.G., Flentke G.R., Lynch M., Kettner C.A.;
 RT Inhibition of Iga1 proteinases from Neisseria gonorrhoeae and
 RT Hemophilus influenzae by peptide prolyl boronic acids *;
 RL J. Biol. Chem. 265:3738-3743(1990).
 CC - FUNCTION: THIS PROTEASE IS SPECIFIC FOR IMMUNOGLOBULIN A.
 CC - CATALYTIC ACTIVITY: CLEAVAGE OF IMMUNOGLOBULIN A MOLECULES AT
 CC CERTAIN PRO-1-XAA BONDS IN THE HINGE REGION. NO SMALL MOLECULE
 CC SUBSTRATES ARE KNOWN.
 CC - SUBCELLULAR LOCATION: SECRETED.
 CC - DOMAIN: THE SIGNAL PEPTIDE GUIDE THE PRECURSOR TO THE PERIPLASMIC
 CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
 CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
 CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS.
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL, X04835; CAA28538.1;
 DR PIR, A26039; A26039.
 DR MEROPS; S06.001;
 DR InterPro; IPR000710;
 DR PRINTS; PRO0921; IGASERPTASE.
 KW Hydrolase; Serine protease; Zymogen; Autocatalytic cleavage;
 KW Transmembrane, Signal.
 FT SIGNAL 1 27
 FT CHAIN 28 986
 FT PROPEP 987 1532 HELPER PEPTIDE.
 FT ACT SITE 278 278 POTENTIAL.
 FT SITE 986 987 CLEAVAGE (AUTO-).
 FT SITE 1018 1019 CLEAVAGE (AUTO-).
 FT SITE 1121 1122 CLEAVAGE (AUTO-).
 SQ SEQUENCE 1532 AA; 168976 MW; 68FF4112BD22440D CRC64;

Query Match 3.3%; Score 176.5; DB 1; Length 1532;

Best Local Similarity 19.18%; Pred. No. 0.91; Indels 437; Gaps 60;

Matches 247; Conservative 157; Mismatches 451; Indels 437; Gaps 60;

16 VLSL-CAYALNQRS-----QENKNNRVSYYDGSQSSQKSENTAPPQ 58
 DB 11 ISLSIFLAYALTPYSEALVADVDYQIFRFAENKKGFFGANDLSVKNRRQNI-GNA 69
 59 VSGKEGI-----QAEQYIKITDQGYTS-----HGDIHYNKGVPIDALFSEEL 105
 DB 70 LSNVPMIDFSVADVNRKRIATVVDPOYAVSVHARAQVHTFYQYNGNDVADENEYRV 129
 106 MKDNYQ-----LKDADI-----VNEVK-----GGYIIRVD-----131

Db 130 VEONNTEPHKAWGASNLGRLDYNNANPNKFEVAPLAPDAGGLDTYKDKNRFSSFV 189
 QY 132 ---GKYVYVKKDAHAD-----NVRTDEI--- 153
 Db 190 RIGGRQLYVEKGYHOBEGNEKGYDLADLSQAYYALAGPYKINDQNTNTEGLIGFG 249
 QY 154 NRQROEHVKDNEKYSNVAVA-----RSQ-----GRYTTNDGY----- 186
 Db 250 NNNROYSAEELKQALSDQDLNTEGLDSDSPLEAFKQKQNWQFGLTYDYMAQYKXW 309
 QY 187 ---VFP---ADIEDGNAYIPHGCHYIPKSDLSASELAARAKHLAGKNNQPSOLS 240
 Db 310 QEWNIYKEEPADKIKODNAGVNGEHHW---KTGTNSHIGSTAVRLA--NNEGDLNNG 366
 QY 241 YSSASDNT-----OSVAKS-----TSRPAKSENLQSLKELXDSPOKRSSES 287
 Db 367 QNTYFENGTLVNLQNIQAGGLEFFKGDYTVKAGN---NDITWLAGID----- 413
 QY 288 DGLVDPKAKIISPTPNQVALPHGDHYHFIPIYSKLSALEKTIAR--WPIISCTG----- 338
 Db 414 ---VADGRKYVQYK---PBGDRL-----AKIGKGLEINGTGVNOGLKV 454
 QY 339 ---STVSINAKPNE-----VVSLSLSNPSLSLTSKELASASDG--YIFNPKD 383
 Db 455 GDGTVILNQADADKRYOAFSOGVIVSGRGTLYLNSNOINPDNLYFGFGRGLDANGND 514
 QY 384 IVEE---TATAYIVHG--DHFHYIPKSNQIGOPTLPNSLATPPSLPINTGSHK 437
 Db 515 LTFEHIRNVEGARIVNHNTHASTI-----TLTGKSLITNNSLSVH--STONDY 563
 QY 438 EEDGYC---PDANRIADESGFVNSH-----GDHNYFFKDLTE 475
 Db 564 DEDDYSTYPRPRPIPOGKLYKRYALKSGRLNAMPENGVAENNDAJFMYGTQE 623
 QY 476 EOLKAQKHL-----EYKTSNGLDLSLSHEDYPGNAKEKDLDKKIEEK 522
 Db 624 EAKKMMNNRNRIDEGFEGFEDENGKGNALNL-----NFGKSAQRFL----- 671
 QY 523 IAGIMQYVKRESIYVNEKNAIIPPHGDHHDADIDEKHPVIGSHSNEL----- 576
 Db 672 LRG-----GANLNGKISVQGNVLSGRPTPHARDFYVKSARKDAHFSKNEVYFEDW 726
 QY 577 ---FPEE-----GVAKKEGKV-----YTGEL 597
 Db 727 INRTFAAELAVQASAFSSGRVSDITANTATDNKAVNLGKNGDEVCVSDYGYT 786
 QY 598 TNYVILLKSTFNQNTLANQOKRVSFPPELEKLGIMLVKLIITPDGKYLEVSGK 657
 Db 787 CNTGN--LSDKALNSFATRNG-----NVMNLQNAALVL--GRA--ALMGK 827
 QY 658 VFEGGVGNANFELDOPYLPQOTFKYTIASKD-----YFVSYD 696
 Db 828 IGGGNSRYSLNQSHKMLTGDSQVHNLSLASHIHLNNSAQASANKYHTIKINHLGN 887
 QY 697 GTFYVTSIAYMASOTIYPPHADDTYLRNPOFAPV--KGTDL--VRVDEHGNAY 752
 Db 888 GHHYITLDKMLGKRVLAKESASHYQLHONKTGEPMQEGLDLFDASSVQDRSLFVS 947
 QY 753 LKNY--KVEILPIPKLNOGTRT---AGKRIYTFMANAYLNOSTIYVFPILKE 807
 Db 948 LNHVYDLAGALNTYI--KTENGITRLYNPYAGNGRPVKKPAPSPAANTASQ-----AQKA 999
 QY 808 NOTD-----KPSIL-----POFRKNAQENSKLDEKVEEPTSEKYEKELSETGSTNS 858
 Db 1000 TQTDGAQIAKPNQIVVAPPSPQANQAEELRQQAALAEQYKROAAAEAEKVAKQDEAKR 1059
 QY 859 TLEEVYVDPVOEKYAKFAESTGKMLENVLFNMGTIELYLPSEGVYKKNMADFTGEAPQ 918
 Db 1060 KAAELIAR---QOEAKKAA-----ELAANKAE----- 1084
 QY 919 GNGEKPSENGVSCTYENQPTENKPADSLPEAPNEKPYKPENSTD--NGMLNPEGVNG 976
 Db 1085 --AEKABELAKQKAEASHQ--ANAKPKRRRRRAILPRPAPVFSIDDYDAKDNSESSIG 1141

QY 977 SDPULDPAL-----EAPAVDPQOKLEK 1000
 Db 1142 NLAHVIPRQGRLEINDYEELP--LEELDEAE 1172

Search completed: September 26, 2001, 22:34:29
 Job time: 233 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 22:29:21 : Search time 73.11 Seconds
(without alignments)
1880.248 Million cell updates/sec

Title: US-09-471-255-2

Perfect score: 5206
Sence: 1 MKRSKTIAGSNAVIVSL.....TELALPGEVAKNLSDFIA 1039

Scoring table: BLOSUM62
Gapop 10.0 / Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database: SPREMBL16.*

1: SP_Archaea.*
2: SP_Bacteria.*
3: SP_Fungi.*
4: SP_Human.*
5: SP_Invertebrate.*
6: SP_Mammal.*
7: SP_mhc.*
8: SP_Organelle.*
9: SP_Phage.*
10: SP_Plant.*
11: SP_Rodent.*
12: SP_Unclassified.*
13: SP_Vertebrate.*
14: SP_Virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 921.5 | 17.0 | 822 | 2 | 09ZHG7 streptococ |
| 2 | 230 | 4.3 | 2485 | 5 | 096134 plasmodiu |
| 3 | 226.5 | 4.2 | 1139 | 5 | 097237 plasmodiu |
| 4 | 223 | 4.1 | 1964 | 2 | 059947 streptococ |
| 5 | 218 | 4.0 | 1873 | 2 | 0924N7 enterococc |
| 6 | 215 | 4.0 | 1694 | 5 | 09NXX1 plasmodiu |
| 7 | 214 | 4.0 | 1694 | 5 | 09T7T5 plasmodiu |
| 8 | 214 | 4.0 | 1733 | 2 | 09K114 staphylococ |
| 9 | 214 | 4.0 | 1733 | 2 | 09K114 staphylococ |
| 10 | 211 | 3.9 | 2647 | 5 | 09UXX0 plasmodiu |
| 11 | 210 | 3.9 | 1704 | 5 | 09T7T4 plasmodiu |
| 12 | 205.5 | 3.8 | 1720 | 5 | 025922 plasmodiu |
| 13 | 198.5 | 3.7 | 1693 | 2 | 09LBB3 streptococ |
| 14 | 198 | 3.7 | 1698 | 2 | 09LCO0 staphylococ |
| 15 | 196.5 | 3.6 | 881 | 2 | 09R3J4 streptococ |
| 16 | 196.5 | 3.6 | 1271 | 5 | 025860 plasmodiu |
| 17 | 196.5 | 3.6 | 2910 | 10 | 09FND5 aradidopsi |
| 18 | 196 | 3.6 | 5005 | 2 | 09P8Z5 ureaplasma |
| 19 | 195.5 | 3.6 | 1927 | 2 | 054875 streptococ |

| | | | | | |
|----|-------|-----|------|----|--------------------|
| 20 | 193 | 3.6 | 3724 | 5 | 077320 plasmodiu |
| 21 | 191.5 | 3.5 | 1236 | 5 | 09GTX2 plasmodiu |
| 22 | 190.5 | 3.5 | 4667 | 5 | 09TV19 caenorhabd |
| 23 | 189.5 | 3.5 | 3257 | 5 | 09V736 dirosophila |
| 24 | 189 | 3.5 | 1826 | 10 | 09FM57 aradidopsi |
| 25 | 187.5 | 3.5 | 2151 | 5 | 09NG79 trichomonas |
| 26 | 187 | 3.5 | 1080 | 2 | 048579 listeria iv |
| 27 | 187 | 3.5 | 4688 | 2 | 09P008 ureaplasma |
| 28 | 186 | 3.4 | 1021 | 2 | 048581 listeria iv |
| 29 | 186 | 3.4 | 1231 | 2 | P73362 streptococ |
| 30 | 185.5 | 3.4 | 1071 | 10 | 09M0D9 aradidopsi |
| 31 | 185 | 3.4 | 2166 | 5 | 09V6Y5 dirosophila |
| 32 | 185 | 3.4 | 2195 | 3 | 002822 saccharomyc |
| 33 | 184.5 | 3.4 | 854 | 2 | 09S3P8 streptococ |
| 34 | 184 | 3.4 | 872 | 2 | 09S4J3 streptococ |
| 35 | 184 | 3.4 | 1591 | 11 | P97868 mus musculu |
| 36 | 184 | 3.4 | 2394 | 6 | 077610 bos taurus |
| 37 | 184 | 3.4 | 3381 | 6 | 077609 bos taurus |
| 38 | 183.5 | 3.4 | 2478 | 2 | 09LCH2 staphylococ |
| 39 | 183 | 3.4 | 1260 | 2 | 09XDB6 streptococ |
| 40 | 182.5 | 3.4 | 1023 | 10 | 09LX42 aradidopsi |
| 41 | 182 | 3.4 | 1946 | 5 | 097291 plasmodiu |
| 42 | 181 | 3.3 | 1979 | 11 | 088196 mus musculu |
| 43 | 180.5 | 3.3 | 2478 | 2 | 09RL69 staphylococ |
| 44 | 180.5 | 3.3 | 5170 | 5 | 017490 caenorhabd |
| 45 | 180.5 | 3.3 | 6994 | 5 | 017343 caenorhabd |

ALIGNMENTS

| RESULT | ID | PRELIMINARY: | PRT: | 822 AA. |
|--------|---|--------------|------|---------|
| 09ZHG7 | 09ZHG7 | 09ZHG7 | | |
| AC | 01-MAY-1999 (TREMBLrel. 10, Created) | | | |
| DT | 01-MAY-1999 (TREMBLrel. 10, Last sequence update) | | | |
| DT | 01-MAY-2000 (TREMBLrel. 13, Last annotation update) | | | |
| DE | HPOTHETICAL 92.4 KDA PROTEIN. | | | |
| OS | Streptococcus agalactiae. | | | |
| OC | Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; | | | |
| OC | Streptococcus. | | | |
| OX | NCBI_TaxID=1311; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN-R268; | | | |
| RX | MEDLINE-99115568; PubMed-9916102; | | | |
| RA | Spellerberg B., Rozdzinski E., Martin S., Weber-Heymann J., | | | |
| RT | Schultzler N., Luetticken R., Podbielski A., | | | |
| RT | "Lmb, a protein with similarities to the lat adhesin family, mediates | | | |
| RT | attachment of Streptococcus agalactiae to human laminin." | | | |
| RL | Infect. Immun. 67:871-878(1999). | | | |
| DR | EMBL; AF062533; AAD13797.1; | | | |
| KW | Hypothetical protein. | | | |
| SO | SEQUENCE 822 AA; 92385 MW; 80E4DEF313481F98 CRC64; | | | |

Query Match 17.0%; Score 921.5; DB 2; Length 822;
Best local Similarity 26.9%; Pred. No. 1.2e-42;
Matches 267; Conservative 152; Mismatches 305; Indels 267; Gaps 31;

| | |
|----|--|
| QY | 1 MKRSKTIAGSNAVIVSLSCAYALNHRQENKDNRRVSVYDGSOSQSKS--ENLRPQ 58 |
| DB | 1 MKRTYGTGVAAILLATHIGSYQLGHNHGLATKQNLIVYDSSKQVAPRTNKTMDQ 60 |
| QY | 59 VSOKEGIAEOIYKIKTDGYSYSHGSHYHYNGKVPYDIAFESEELMLKPNYOLKDAI 118 |
| DB | 61 ISAEBSIAEQIYKIKTDGYSYSHGSHYHYNGKVPYDIAFESEELMLKPNYHFRQSDV 120 |
| QY | 119 VNEVKGYYIKVDGKYYVYKDAADNVTKDEINRQKQEHYKD-NEKYNNAVA----- 172 |
| DB | 121 INELDGYVAKNGNYYVYVYKPSKRKNIRTKQOIAEQVAKGKAEKGLAQVAHLISKE 180 |

Query Match 4.3% Score 230; DB 5; Length 2485;
Best Local Similarity 18.0% Pred. No. 0.00074;
Matches 209; Conservative 154; Mismatches 397; Indels 400; Gaps 47;

RA Koonin E.V., Shalimov S., Mason T., Yu K., Fujii C., Pedersen J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA Salzberg, S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.,
RA "Chromosome 2 sequence of the human malaria parasite Plasmodium
falciparum".
RL Science 282:1126-1132(1998).
RL -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMIL; AE001376; AAC71820.1; -
DR HSSP; P24941; ICKP.
DR InterPro; IPR000719; -
DR InterPro; IPR002290; -
DR Pfam; PF00069; PKinase; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR SMART; SM00230; S_TKC; 1.
DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
KW KW
SQ SEQUENCE 2485 AA; 293766 MW; 4297C5CF030AD7E CR664;

DR HSSP; P24941; ICKP.
DR InterPro; IPR000719; -
DR InterPro; IPR002290; -
DR Pfam; PF00069; PKinase; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR SMART; SM00230; S_TKC; 1.
DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
KW KW
SQ SEQUENCE 2485 AA; 293766 MW; 4297C5CF030AD7E CR664;

```

Db 1181 NMLELEYDNCLAOKEENIFRPLNLRKKDVMKRFNKNKIKTIHNEKKRIYQTN 1240
Qy 606 NSTFNQONT-----LANGQKRVSEFPPELEKKIGINMLVLTTPDGKVLKRGVKG 660
Db 1241 KVFEPYNNRREYNELIN---HLTYNFR-----NDLEFL-----SYKSM 1278
Qy 661 EGVGN--INAELODPY-----LPGO---TFRTYIASKDPE---VSDGTFTVPSLAY 707
Db 1279 NNIIRILYINANKIRINNNYDMKLYNOMITLTYQVAANDIDNDHICKGGGLDYIMNITSK 1338
Qy 708 KASQTIFFPAGTGYLEVNPOFAVPKGTDLAVRFDEFHNAVLT-----ENNY 757
Db 1339 ECKNR-----KQKTYL--KKIFHYKKKKDARFFINDEIGSNDMDYDKKYSNDENNY 1389
Qy 758 KVGELKLPKLNQGTTRFAGNKKIVTFMANAYLDNOSTIYVPILEKQNDKPSIIP 817
Db 1390 KLINE-----KMNISMSNDEMDI---PLINSEHONNPPSCOP 1422
Qy 818 CPKRKAGQNSKLDKEVEPKTSEVKEKEKLETSQNSNSLSLEAPYDPVQEKYAKRA 877
Db 1423 NILEKK---STYIDLNLDSNSMDPTEKRYFNVEN-----DLFNTKRMFN 1468
Qy 878 ESYGMKLEAVLF-----NMDGTIELYLPSEGVYKKNMAFTG-----EAPQNGEN 923
Db 1469 FPKGKRLFPNNKFPNNSNEDGVSEF-----KNNILFRELKSNNSLKLESYKNSNN 1520
Qy 924 KPSENGKYSTGVENOPTEN 943
Db 1521 CSNNKGDDNIGNMENNTTN 1540

RESULT 3
097237 PRELIMINARY; PRT: 1139 AA.
ID 097237;
AC 097237;
DT 01-MAY-1999 (TREMblrel, 10, Created)
DT 01-OCT-2000 (TREMblrel, 15, Last sequence update)
DT 01-MAR-2001 (TREMblrel, 16, Last annotation update)
DE HYPOTHETICAL PROTEIN, PFC0235W.
OS Plasmodium falciparum.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=5833;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=93376085; PubMed=10448855;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Felltham T.,
RA Genies S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagals K., Jassal B., Kyes S., Mclean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajadaram M.A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT falciparum."
RL Nature 400:533-538(1999).
DR EMBL, AL034558; CAB39003.2;
KW Hypothetical protein.
SQ SEQUENCE 1139 AA; 135723 MW; 761530047DA11FF9 CRC64;

Query Match 4.2%; Score 226.5; DB 5; Length 1139;
Best Local Similarity 18.2%; Pred. NO. 0.00035;
Matches 216; Conservative 195; Mismatches 455; Indels 333; Gaps 54;
Qy 33 NEDNNRVSY--VDGSQSQSKSENTLPDQVDSQKEGIAEQYIKITDQGYTSHGDHYH 89
Db 62 SKLDNNISFTYTTNGOVEQENE-----IKKQNLHLHQI-----YIDNDEKRYK 107
Qy 90 YNGKYPYALAFEBELAMDPNQQLADADLVNVEKGYIIKVDGKTYVYLKQAAHADNRYT 149
Db 108 GNSNMLHKN--EEEHKKGSPNNIYEID-----NNLEKKKTYNSDKHC--IDE 153

```

[illegible]

[illegible]

```

Oy 704 SLAYMAASQITFYPPHADDTYLKRVPO-----FAVKGKIDALVYRDEEHGNAKYLNNKV 759
Db 444 AVV-----TDKGET--EYQPESPDVTYVDKSGPEQVAPLPEKGN-----I 482
Oy 760 GEIK-LPIPKL-NGGTRTAGNKIPYTFMAAYLDNOSTYVVEYPIIEKENQDCKPIL 816
Db 483 EQVRETEYVEKTEKGGPEKT--EETPV-----KPLEELFPVNPBETTGSTGSI- 527
Oy 817 PQFKIN---KQENSKLDEKVEEPTSEKV-----EKELISETGNSTS 856
Db 528 -QEAENPVQPAEESTNSEKVSPTDISENTGSEVSSNPDSSTSVGESNKPBEHNDKRNES 586
Oy 857 NSTLEEYVTVDPVQAKVAKFAESYGMKLENTLFINNDGTLELPLSGEVIYKKNMAJPTGFA 916
Db 587 EKTVEEVP-VNP-----NEGIVE-----GTSNQETEK 612
Oy 917 PQGNENKPESEKGV---STGVENOPTENKP--ADSLPEAPNEKVPVENS---TDNG 967
Db 613 PVQAPLEEVQTNNGKLANENTGEVSNKPSDSKRPVESHNPENKGNATRPENSGNTSENG 672
Oy 968 KLNPE 972
Db 673 QTEPE 677

RESULT 5
0924N7 PRELIMINARY; PRT: 1873 AA.
ID 0924N7
AC 0924N7;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE SURFACE PROTEIN PRECURSOR.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MH5594;
RX MEDLINE=99081742; Pubmed=9864215;
RA Shankar V., Baghdadyan A.S., Huycke M.M., Lindahl G., Gilmore M.S.;
RT "Infection-derived Enterococcus faecalis strains are enriched in esp,
RL a gene encoding a novel surface protein."
RL Infect. Immun. 67:193-200(1999).
DR EMBL, AF034779; AAD09858.1; -.
DR InterPro; IPR001899; -.
DR Pfam; PF00746; Gram_pos_anchor; 1.
KV SIGNAL.
FT SIGNAL.
FT CHAIN 1 49 POTENTIAL.
FT CHAIN 50 1873 SURFACE PROTEIN.
SQ SEQUENCE 1873 AA; 202084 MM; F609483DB980621 CRC64;

Query Match 4.0%; Score 218; DB 2; Length 1873;
Best Local Similarity 20.8%; Pred. No. 0.0022;
Matches 215; Conservative 122; Mismatches 377; Indels 322; Gaps 49;
Oy 91 NGKVVYDALFSEELMKRPNTQAKADIVNEKGGYITIKVQCKYVYLKDAHAD-----N 146
Db 593 NKNLEDEAEYSKTEPTDSNTDSTKGIVYAKIGRTEDVDVEFVAKSAQENENATYVP 652
Oy 147 VTKDEINRQCKEYKDNDEKVNANVAARASQGRYTTNDGY-----YFNPAIIETDGNAYI 202
Db 653 ITTPETITIQSKRPFPDKDPV-----LANDAFSLVDIYN-----KDSGNMSV 695
Oy 203 VPHGHYHYIPKSDLSASELAAAKAHLAGKNQPSQLSYSTASDNNNTQSAKGSTSKPA 262
Db 696 DANNGIVTFPAKVGGESEPIGTI-----PIKIYV-----QDSGVGTDLAV 738
Oy 263 NKSEKLSGLLEKIDSPSAQ-----RYSSEGLVYDDPAKIIKRTNGVAILPHGDHYH 314

```

Db 739 TVSKN-----IYENGENIPAGYHKTFTAGEGTSIESGTTVEAVKGVSLPE----- 786
 QY 315 FIPYKSLAEKIRAMPISGTSVSTINAR-PHEVYSSLSGSSNSPSLSLTSELSA 373
 Db 787 ----DKLPYKAKNDG-----YTDAMPEB-----ATOPKADDFEVSSA 822
 QY 374 S--DGYIFPKX-----IYETATAYVRGHDHNYIPKSNQIGOPTL 414
 Db 823 YKLDIIEPNDINPAGIKYKFTFTAGEGTSIESGTTVEAVKGV-----VSL 868
 QY 415 PNNSLATPSPLPIMPSTSHKHEEDGYGFDAKRIIAEDESQVNS-----HGDD 464
 Db 869 PEDKL---PYKADGVYDAMPE-----EATOPKADDFEVSSATKLDIIEPNDIN 919
 QY 465 -----NHYFFKQD---LTEBQ---IAAKHLE----- 486
 Db 920 IPAGYHKTFTAGEGTSIESGTTVEAVKGVSLPEBKLPYKAKDGYTDAMPGEATOP 979
 QY 487 -----EYKSHNGDLSISHEODYPQNAKEMDLOKRIEKLAKIKGYKRESTITVNR 541
 Db 980 KADDFEVSSATKLDKSDADKTPBGOVYTELKKE-PDASEGINKKDLKPKADKTYTK 1038
 QY 542 E-----KNATYPRGHDHNDPIDEHNPVIGISHSHNYELKPE-EGVAKKEGK 590
 Db 1039 EKVDTISAGNKGTVVYVTS-D-GSSDEVE--VDVYTDNRSDADKTEPVEGEKVEIGK 1095
 QY 591 VTTGEELIVNNILKNSTNN-----ONFTLANGQKAVSSPPELEKIKGINLVYL 643
 Db 1096 VDLTDVNTMLPPLPGQTYTDTVPQGTIDTNPNGEVIEVTPD-----GKTDVTV 1149
 QY 644 ----IPPDGVLEKVSQKVEGEGV-----GNANFELDQYLPQGTQKTYTASQDVP 691
 Db 1150 PVEYVDNRSDADKTEPVEGEKVEIGKVDLDNVTNL-----PTLPQGT---TVT----- 1197
 QY 692 EYSDYDGTFTVPSLAKMASQTIFFPHAGDYLTVRNQFAPVPGKTDALVRVDEPHGNA 751
 Db 1198 DVTGGTIDTNPNGEVIEV-----TY-----PDGKQDVKVPVEVTDNR 1239
 QY 752 YLENNY-----KV-----GEIKLP-----IPKINQGT---RTAGKIKIYVTRAAKAYILDN 793
 Db 1240 SDADKTEPVEGEKVEIGKVDLDNVTNLPLPQGTYYTDTVPQGTIDTNPNGEVIEV 1299
 QY 794 GSY-----IYEVPLKEKNOTDKPSILPFRKNAKQNSKLDKVEEBKSEKYEK 847
 Db 1300 EYTPDGTDTVKAPEVETDNRSDADKTPMVEGEKVEIGKVD----- 1343
 QY 848 LSTGSGTSTNSLLEFPVDPVQEKVAKFAESYGMKLENVLFNMGTIELYLPSSG---EVI 905
 Db 1344 --LTDVNTMLPPLPGQTYTDTVPG-----GTIDTNPNGEVIEVTPDQKQDV 1393
 QY 906 K-----KNADFTGAPGNGENKSENGKYS--TGTVENOPLENKPADSLPEAPKEKY 958
 Db 1394 KYPVEVTDNRSDADKTEPVEGE-KVEIGKVDLDNVTNLPL-----LPQGTVTV 1445
 QY 959 KPESTDNGLNLEGN 974
 Db 1446 TPEGCTIDT---NTPGN 1458

RC STRAIN-FCCL/HN;
 RA Shan 2.X., Yu X.B., Li X.R., Ma C.L., Fang J.M.;
 RT "Molecular cloning and sequence analysis of major merozoite surface
 RL antigen(gp195)gene of Plasmodium falciparum isolate FCCL/HN."
 DR EMBL; AF218248; AAF27536.1;
 DR InterPro; IPR000561;
 DR InterPro; IPR003247;
 DR Pfam; PF00008; EGF; 1.
 DR Pfam; PF000527; -, 1.
 DR Merzoite.
 SQ SEQUENCE 1694 AA; 192766 MW; B51634449E0F6728 CRC64;
 Query Match 4.0%; Score 215; DB 5; Length 1694;
 Best Local Similarity 20.3%; Pred. No. 0.0027;
 Matches 248; Conservative 181; Mismatches 409; Indels 384; Gaps 67;
 QY 1 MAFSKRYIAGSAVIVSLSLCAVALNQHRSQENKDNRRVSYVDS---QSSQKSENTLP 56
 Db 500 MAFNNF---DKDVIKIFSAITYN---VEKQYNNKSSSSNNSVYVQKLLKALSTYLE 553
 QY 57 DOVSQKEGIAQIYIKITDQGYTSHGDHITYNGKPYDA---LFSREL----- 104
 Db 554 D-YSLRKGJ-----SEKDPNHYTLKTLGLADIKRLKEIKSSENKJLE 596
 QY 105 -----LMDPNYQKADADIVN-EVAGGYII-KVDS--KYVYVILKDAHADNVRT----- 149
 Db 597 KNEKGLTHSANSLESDVVKIQVAVTLIKKIEDIKLELKLKNAQLKDSIHVPIITYP 656
 QY 150 -----KDEINRQK-----EHVKDNEKYNVA---VARSQGRYTDNG--- 185
 Db 657 QNKPEPIYIYVAKKEDKKEIPKVKMLKKEQAVLSITQPLVAAS---TTEDEGHS 713
 QY 186 -YVFPADIIEDTGNAYIVPGCHYHI-----PSSLSASELAAKAKLAKNNQPSOL 239
 Db 714 THTLSQSGTEETEEVY-----GHYTYVITLPTPOPSPPEKVEKVENSIERK 761
 QY 240 SYSSASDNNTNSVAKGTSKPAK-----SNNLSLKELYD-SPSAOR 283
 Db 762 -----SNNSQLTITVYTKLDEFLTSYCHKIYLVSSNMOKLELVNLRPEKN 815
 QY 284 YESDQVLDPAKIIISRTNGVALIPHG-----DHYHFTPYSK---LSALEEK--- 327
 Db 816 ELKS---CDPLDLEFNQNNIPAKSYLSDSMNDLQHLFFELYQEMLYLHKLEENH 871
 QY 328 ---IARVPISGTSTVST-NAKPREVYSSLSJSSNPSLTSKELSSASDGIYFNK 382
 Db 872 IKKLEQKQITGSTSTSPGNTVTVAQSATHSNQNSQNSAAS---TNTONG----- 922
 QY 383 DIVERATAYIVRHGDHFIYIPKSNQI--GQPLPN--NSLAPSPSLPINTGSHKHEE 439
 Db 923 -VAVSSGPAVVEESHDLTVLISINDLKIVSLNIGNTKVYPNP-LTIST-TENEKFE 979
 QY 440 DGYGPDANRIIAEDESQGFVNSGHDHNYFFKQD-----LTERQIAAKHLEEVTS- 491
 Db 980 N-----ILANDTYF---NDLIKQFVKNKSVYITGLTEGOKNALDEKIKLKDIO 1027
 QY 492 -----HN-----GLDLSISHEODYPGNAKEMD---LDKRIEKLAKIKGYKKE-SIV 538
 Db 1028 LSPFDLYNKYKLLDRLFNKKKEIGQDKQIKKTLTKLEQLESKLSINPHVYLQNFYVF 1087
 QY 539 VAKENKAI-----YPRGDHNDADIDEKIPGI---GSHSNV 574
 Db 1088 FKKKEAEIAETENTLENTKILKHYKGLVKKYNGE---SSPLKLTSEVSIQTEENTYAVL 1144
 QY 575 ELKPEEGYAKKEGNVYVGE-----LNNVYILKNSFNQNFNFTLANGQKRVSS 626
 Db 1145 EKFRYLSKIDGKLNDLHGKKLKSLSSGLHLIYELK-EYIKNNYI-----GN 1194
 QY 627 PPELEKIKGINLVYL---IPPDGKVL-----EKVSKVGEVGNIANFELDQPYL 676
 Db 1195 SPSENKTK--VNEALKSYENFLPEAKVTVTVVPPQDVPSPSLSVYSSSSSTSENOI 1252

QY 677 PGO-----TFRYTIASKDYPEVSYDGTFTVPTSLAYKASOTIFYPFHAGDTYL-RVNPQ 730
 DB 1253 PFGSLTLELOQVOVOSONDEED-DSLVLPL-----IFGESEDNDEYLDQVVTG 1300
 QY 731 FAVPGGTALVAVDEEFGHGNATLENNKAYGEIKLPIPLKNOGTTTACNKTIP---VFMA 787
 DB 1301 EASVTMNLISGF-----ENEYDIYTLK-PL---AGVYRSLKKQLEKNITITFNL 1346
 QY 788 NA-----YLD-----NOSTYIYVPILEKENQTDKPSILPOFRN 822
 DB 1347 NLNDLNSRLKRRKFTFLDYLESDLMQFKHISNEYIIE-DSFKLLNSBOKNTLLSKYKI 1405
 QY 823 K-----AOENSKLDEKY-----EKPTESEYKEREKLTSTGNSSTLEEVPTVPD 868
 DB 1406 KESVENDIKFAQEGSIYEKYLAKYKDDLESIKYIYE---EKERPSSPTPTPPSPAKTD 1462
 QY 869 VOEKAKFA-----ESYGMKLENYLFNM-----891
 DB 1463 EOKKSKFLPFLTNIETLYNMLVNNIDYLLMLKAKINDCNVEKDEAHYKTKLSDLKAI 1522
 QY 892 DGTIELYLPSEGVKKNMADEFGAPOGNGENKPSENGK-VSTGTVENOPT-----EN 943
 DB 1523 DDKIDLF-----KNTNDEFAIKKLLINDTKMDLKGKLLSTGLVQNPPTIISKLEIG 1574
 QY 944 KPADSLPEAPNE--KPVKPENS 963
 DB 1575 KFOQMLNISQHCQVKKOCPENS 1596
 RESULT 7
 ID 091215 PRELIMINARY; PRT; 1694 AA.
 AC 091215;
 DT 01-MAY-2000 (TREMblrel. 13. Created)
 DT 01-MAY-2000 (TREMblrel. 13. Last sequence update)
 DE 01-JUN-2000 (TREMblrel. 14. Last annotation update)
 DE MEMOZOITE SURFACE PROTEIN 1 PRECURSOR.
 SN MSP1.
 CS Plasmodium falciparum.
 XC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium
 NC NCBL_TaxID=5833;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=HN1;
 BL Jiang G., Liu R.Z., Daubenberger C.A., Pluschke G.;
 BL Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
 CR EMBL, AF062348; AAC72884.1;
 CR InterPro; IPR000561;
 CR Pfam; PF00008; GGF; 1.
 KW Merozoite.
 SW SEQUENCE 1694 AA; 192794 MW; 84CECDE/09F5673B CRC64;

QY 186 -YVENPADIIEDTGNAYIVPGHGYHI-----PKSDLSASELAATAKHLAKNNQPSQL 239
 DB 714 THTLQSGSETEETEY-----GHITTYITLIPQPSPPKPVKVEANSIEHK-----761
 QY 240 SYSSASDNNTQSAKSGTSKPAK-----SENQSLLEKELD-SPSQR 283
 DB 762 -----SDNSQALTKRYTLKLDLEFLKSYCHKYYIIVSNMSMOKILEVYNLPPEKN 815
 QY 284 YSESDGLVFPDAKISRPNCAIPIHG-----DHYHFIYK-----LSALEEK--327
 DB 816 ELKS-----CDPLDLEFNQNNIPANYSLYDSNNNDLQHPFLYCKEMYYIHLKLEENH 871
 QY 328 ----IARVPLSGTSTYST-NAKPNYVSSLSGSLSPSSLTSEKLSADGYTFNPK 382
 DB 872 IKLLLEQKQITGISTSPGTTVTNQAASHNSQNOQASAS--TNQNG-----922
 QY 383 DIVEETAJAVYVRGDFHYIYKSNQI-GQPLPN--NSLATPSPSLPINTGSHKHEE 439
 DB 923 -VAVSSGPAVVEESHDPVLVLSISNDLGIYSLNMGKTKVPNP-LTIST-TEMEKETE 979
 QY 440 DGYGDARIIAEDSGVYMSHGDHNYFEKKD-----LREQIKAKOKHLEEVKTS--491
 DB 960 N-----ILKNQTYF-----NDIKQFVNSKRYITGLTETQNALNDEKTKLDTQ 1027
 QY 492 ----HN-----GLDLSSEHODYPNAKEMD--LDKRIEIKINGIKQGVKRE-SIV 538
 DB 1028 LSFQLYNKYKLLKDLFKKKKELGODKQIKKTLTKQLBESKLSLNNPNNVLONFSEF 1087
 QY 539 VKERKNAI-----YPRGDHHDADIDEKPVGI--GHSHSNY 574
 DB 1088 FNKKKEAIEAETENTLENTKILKHKYKGLVAYNGE--SSPLKLTSEVSJOTEDENYANL 1144
 QY 575 ELFKPEEAVAKREKQKVTGEE-----LTNYYNLKNSFTNNQNTLNGKQKRVSE 626
 DB 1145 EKFRYLSKIDGLNDLNLGKKSFLSSGLHLLTEK-EYIKKNT-----GN 1194
 QY 627 FPELEKRLGIMLVKL--YTPDGKVL-----EKYSGRVGEGVGNANFELDQPYL 676
 DB 1195 SPSENNK--VNEALKSYENFLPEAKVTTVTPQPDVTPSLSVRVSQSSGSTREEQI 1252
 QY 677 PGO-----TFRYTIASKDYPEVSYDGTFTVPTSLAYKASOTIFYPFHAGDTYL-RVNPQ 730
 DB 1253 PFGSLTLELOQVOVOSONDEED-DSLVLPL-----IFGESEDNDEYLDQVVTG 1300
 QY 731 FAVPGGTALVAVDEEFGHGNATLENNKAYGEIKLPIPLKNOGTTTACNKTIP---VFMA 787
 DB 1301 EASVTMNLISGF-----ENEYDIYTLK-PL---AGVYRSLKKQLEKNITITFNL 1346
 QY 788 NA-----YLD-----NOSTYIYVPILEKENQTDKPSILPOFRN 822
 DB 1347 NLNDLNSRLKRRKFTFLDYLESDLMQFKHISNEYIIE-DSFKLLNSBOKNTLLSKYKI 1405
 QY 823 K-----AOENSKLDEKY-----EKPTESEYKEREKLTSTGNSSTLEEVPTVPD 868
 DB 1406 KESVENDIKFAQEGSIYEKYLAKYKDDLESIKYIYE---EKERPSSPTPTPPSPAKTD 1462
 QY 869 VOEKAKFA-----ESYGMKLENYLFNM-----891
 DB 1463 EOKKSKFLPFLTNIETLYNMLVNNIDYLLMLKAKINDCNVEKDEAHYKTKLSDLKAI 1522
 QY 892 DGTIELYLPSEGVKKNMADEFGAPOGNGENKPSENGK-VSTGTVENOPT-----EN 943
 DB 1523 DDKIDLF-----KNTNDEFAIKKLLINDTKMDLKGKLLSTGLVQNPPTIISKLEIG 1574
 QY 944 KPADSLPEAPNE--KPVKPENS 963
 DB 1575 KFOQMLNISQHCQVKKOCPENS 1596
 RESULT 8
 ID 091114 PRELIMINARY; PRT; 1733 AA.

AC 09K14
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE PUTATIVE CELL-SURFACE ADHESIN SDRF
 GN SDRF
 OS Staphylococcus epidermidis
 OC Bacteria; Firmicutes; Bacillus/Clostridium group
 OC Bacillus/Staphylococcus group; Staphylococcus
 OC NCBI_TaxID=1282;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9491
 MEDLINE=20340957; PubMed=10878118;
 McCrea R.W., Hartford O., Davis S., Ni Eldin D., Iina G.,
 Speziale P., Foster T.J., Hook M.,
 The serine-aspartate repeat (Sdr) protein family in staphylococcus
 RT epidermidis.
 RL Microbiology 146:1535-1546(2000).
 DR EMBL: AF245041; AF72509.1;
 DR InterPro: IPR001899;
 DR PROSITE: PS00343; GRAK_POS_ANCHORING; UNKNOWN_1
 SQ SEQUENCE 1733 AA; 184720 MW; D8D62EA1692D4E8 CRC64;

Query Match 4.08; Score 214; DB 2; Length 1733;
 Best Local Similarity 18.5%; Pred. No. 0.0032;
 Matches 226; Conservative 175; Mismatches 458; Indels 360; Gaps 55;

QY 2 KFS-KYIANGSAYISLCAVYLNCH-----RSQENDNNRVS 40
 DB 22 KYSRKRTVGTASLIVATLFGAADNEAKAEADNQLSASKEQKSRDNENSKLNQVD 81
 QY 41 YVDSQSSQSKSEMT-TPDOYSQKEGIAEQIVIKITDQGYTSHGDHYHYNGKVPYDAL 99
 DB 82 LDNGSHSEKETTNNATNATVETKVEAPTSDVSKPKANEAIVTN----- 124
 QY 100 FSEELMKADPNYQADDIYNEVKGITIIYVDGYYIYLLDAHADNVKTKDELNRK-- 157
 DB 125 -----ESTKPKTEAPTNEES-----IAETPKSTTQDDSTEKNP-SLKDNINSSSTT 173
 QY 158 QEHYKDNENSNVAVARSGRTYNDGYFNPAADIIEDGNAIYVHGHHYHPIKSDL 217
 DB 174 SKESKTEHSTKQAKQSTNKSNDLTND-----SPQSEKTSQA-----NND5 216
 QY 218 SASLIAAKAHLAGK-----NMQPSQ-----LSYSTASDNNQSVAK 255
 DB 217 TDNSASPK-QLOSKPSEQVYKTKFNFDEPTQDVEHTTKLPKPSVSTDSVNDKQDYTR 275
 QY 256 GSTSK---PANKSENLOSLLKELYDSPAQRITSESGLVFDPAKTISRTPN-GVAIP-- 308
 DB 276 SAVASLGVDSNEFEATINAVRDNDLPAASREQINEAIIAALKKDSNDPDGYDTFLAL 335
 QY 309 -----HGDIYHFIYPSKLSALEEKIARMPISG-----TGSTVSTNAKNEVYSS 353
 DB 336 NRSQSKSPKASAPRANSLAAE-----PNSGKNVNDKVKITNPILSLKSNHANNV 390
 QY 354 LGSLSNPSSILTSKEL-SSASDQYIPNR----- 382
 DB 391 IMPTSNQPIKANKYELDDSIKEDGFTIKYGYIRPGLELPAIKQQLSKDSIYANG 450
 QY 383 --DIVEETAAIYRBDHRIYIPKSNQIGOPILPNNSLTPSPSLPINTGSHKEEED 440
 DB 451 VYDVTYTTTYTFTNYDOYQONTIGSFDL-----IATPKRETAIK-----DNQ 493
 QY 441 GYGRD---NARIIAEDSGVSMGHNDHNFKKDLDEEDIKAAQKLEVEKYSHNGDS 497
 DB 494 NTPREVTIANEVAKD--FTVDGN-----KQDNT--TTAAVAVANDNNKHNHNVY 541
 QY 498 LSHEDD-----YPG-----NAKEKQDJKREE 521
 DB 542 LNNQNNPKYAKYSTYKNGEIPGEVYEVVDITNANVDSFNDLSSNVKQDVTSGAP 601

QY 522 KIGIKQYGVKRESIYVNEKNAIYIPGHDDHADPIDEKKFVGIGSHSNYLEKPEE 581
 DB 602 KYASDGRDINDINARSMANKKIY-----TQVRRTGNGNYTEWLR--D 647
 QY 582 GVAKKGNKRYTGEELTNVNLKNSTFNQNTLANQ-----KVSFSPPELEKKGI 637
 DB 648 GTNT--NDRYRGKSTVYLLNGSSTAQGDNPYLSIGDVMIDKKNNGVODDEKGLA- 704
 QY 638 NMVLKLTTPDGKYLEKSGVPEE-GVGNIAN-----FELDDPYLPQGFRTKTSKOV 690
 DB 705 GYVYTLKDSNNRELQRYTDSQSHYQFDNLQNGTYVEFAIPDNYTPSPANNSTDAID- 763
 QY 691 PEVSDGTFTVPTSL-----AYKASQITFY-----PEHAGDTYLRNVPQFVPGTALVR 742
 DB 764 SDEBDGTNRVYVAKGTINNADNMTVDGTFLPKTINVD-----Y 804
 QY 743 VPEDFGNALENNYK-VGEIKLPILPLNO--GTRTAGN-KIPVTFMANAYLDNSTY 797
 DB 805 VMEDTNKDGIQDDNENGISGVKTYLKNKNGDTIGTTTDSNGKYEFTGLBN-----GDY 858
 QY 798 IVEVPILEKNOQDKSILPQFRN-----KAQENKLDKVEEPEKTS-----EK 842
 DB 859 TIEFETPEGTPTKONGSDGKDSNGTKTIVYKADNRTIDSGFYKPTYNAGDYWED 918
 QY 843 VEREKL---SETGNSTNSYLEEPTVDYQEKVAKFAESYGM-K-LENYFNNQGTIELY 898
 DB 919 TNDGQIODESEKISGVKTYLKD-KGNALIGTTTDSAGHYQKGLENGSY-----TYEE 973
 QY 899 LPGEVYIKMMADFTGAPQNGENKPSKNGKYSTGTVENQPTENKPADSLPEAPNEKPV 958
 DB 974 TPGS-----YTPKANSQGDITVDSNGITTTGLING-----ADNL----- 1008
 QY 959 KPNSTDNQML-NPEGVSGDPMIDPLAEAPVADPYQELKFTASYGIGDSYIFNM 1017
 DB 1009 -----TIDSGFYKTPKYSVG-----DYWEDTNK-DGIQDNEK-----GISGV----- 1046
 QY 1018 GTIELRPSGEVYIKNLS 1036
 DB 1047 -KYLKDEKGNITSTTTD 1064

RESULT 9
 Q25662
 ID Q25662 PRELIMINARY; PRT; 1939 AA.
 AC Q25662;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)
 DE REPEAT ORGANELLAR PROTEIN.
 OS Plasmodium chabaudi.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5825;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=96V;
 RA Weiner E.B., Taylor W.R., Holder A.A.;
 RL Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.
 DR EMBL: U43145; AAC63403.1;
 SQ SEQUENCE 1939 AA; 229001 MW; B36E462001C6F22F CRC64;

Query Match 4.08; Score 214; DB 5; Length 1939;
 Best Local Similarity 17.9%; Pred. No. 0.0038;
 Matches 220; Conservative 183; Mismatches 415; Indels 408; Gaps 53;

QY 28 HROENDNNRVSIVDSQSSQSKSEMTLPDOYSQKEGIAEQIVIKITDQGYTSHGDHY 87
 DB 462 HKELNELNQLSKLNKKNINEN-----TELNDKI 494
 QY 88 HYNGKVPYDALFSEELMDPNYQLKA---DIVNEVKGITIIYVDGYYIYLLDAAH 143
 DB 495 SSLNSEY-----NILNDKQTLGNDITLNDLNNLR----- 526

```

144 ADVERTKD-EINROGAEHVKDNEKYNVAARSGRTTNDGYF-----NPADIT 194
527 -NEIWTSDKMKKMEEDLAMEEKEGCVUIDEIEKYNKIFLEIEKKEKRYADLN 585
195 EBTG---NAYTPRGHGHYIPKSDLSASELAALAGKMGQSPQSLSTSTAD----- 247
566 DEISILRNSIYKE-----KEPIEKEFEYENKINLENNKEEKNKIYENELSLRK 637
248 -NNTOSVAKSGTSFKPANKSENOSLKEKY-----DSPSQRSESDGYF 292
638 YDNEGGLKQIDELMIOKLTBEKYL-QIYDNNMMPRSICKIDMPSITKSGDLYDF 696
293 DPAKIIISIPNGV-AIPGDHIFIPYSK-----LSALEEKIAPVPISTGT----- 338
697 VYAYIKRDESSDANPOTHEKEMVALEEKRAALVAALEEKHEKELAKGCHGVYLR 756
339 -----STVSINAPRNEVYSLG-----SLSSNPSLSTTS----- 367
757 LGEOKKEBTIIEEKHKVDYTKLGEOKHENTIKLEEKHDVYTKIGDQYKEIATKEEH 816
368 -----KELSSASGYIFNPDIYEETATAYVRGDFHFIIPKSNIOGPTLNN----- 416
817 AVYVALEEKHKLGSCH-----KAWDELEK-----RHAUPEGLEEKHKEATKLEGHKS 868
417 -----NSIAPPSBPLPINPSTSH-----EKHEEDGYFDMRIIAEDS 455
869 ENNEVEKHADPEVELEEKHAEFTAKLGEHREYVAGLEEKKE-----VYALEE 919
456 GFVMSGHGHNHYFFKDLTEQIKA--AKHLEEKYTSNGIDSLSSHBODYPGNAKEMK 513
920 -----HKKEIALIEGHEKEMVALEEKHEVYA-----GLEAKHNLEEGKEVAALE 968
514 -----DLDKKIEEK-----IAGIMKOYGVRSIYVNNKKNMIIYHGH 554
969 KRAHDLVAVLEBQHAEIITKIGEEKEVYAGIEEKY-KVEATKLAEBKVDVYKLGSH 1026
555 H--ADPIDEKPV-----GIGSHSN-----YELFKPEE 581
1027 KEELAKLEDGHEVYNEYEKKNASILLNLEENHKEMIKLEHEHESASDLVEKLYODE 1086
582 GYAKREKGVYTGELTAVNNLAKSTNNONFTLANOGKRVSEFPPELEK-KIGIML 640
1087 EV-KNSNKKI-----EELTAVIDLNDISICYKQILEEYERNEYV--BEINKLKIVONEM 1140
641 VKLITPDGVLEKVSIG-VYFGGVGINANFELDOPYLPQFEKVIYIASKDPEVSYDGF 699
1141 KDM--NOKKIIIEKEIKKLNKLSNIVETKE-----NTYK----- 1176
700 TVPTSLAYKMASQITFYFPHAGDYTLRVNPOFAVPEKGTDALV--RVDFEFGNATLENNY 757
1177 -----NSEMVYNNENKRIIIVOSKENISSESVEB-- 1206
758 KVGEEK--LPIPKLNOGTRTAGNKIPTYPMANAYLNOSTIYVAVPLEKNOGDKRSI 815
1207 KGNLKMTLSTLKKERNIFSINDKNESSELVDYI--KSAYINKIEYKKEIEIDGNKI 1263
816 -----LPOFRKNAQOENSKL-----DEVEEPEKTSKEK----- 844
1264 EDLKNKIIDLNELLINENKKNVLTDENNMLKKEIETDKNLNKEKNENBEILANDDI 1323
845 ---KERLSETGNSNSTLEEVPTVDVPOEYKAFASGYKLENTL--FN-----M 891
1334 IKLKKEISEMDEEKEKLTKENI-----KLKNDIEQINKEIKKEELMKLFENINEVYSL 1379
892 DGTIELYIPSEVYKKNADPTGEAPONGENKSEKSVSTGIVENOPTKKNRADSLE 951
1380 KNOIEKMKLEELKNKNTELLAEKRETNMSISNDKNIVNNITL--TSK----- 1430
952 APNEKPVPEPSTNGMLNPGVNSDPLDPALEAPAVDPOEKEKLEKFASTYAGL-- 1009
1431 ---QNNLNR-----NVEDKTGDDINCEKNNDQAEISYLDDEIKKISMLTGEELNR 1478
1010 ---DSVIFNNDG---TIELRLPSGE 1028

```

```

Db 1479 KNSYDEKYNLNLKELKRNKKE 1504
RESULT 10
ID 0904X0 PRELIMINARY; PRT; 2647 AA.
AC 0904X0;
DE 01-MAY-2000 (TREMBLREL 13, Created)
DE 01-MAY-2000 (TREMBLREL 13, Last sequence update)
DE 01-MAR-2001 (TREMBLREL 16, Last annotation update)
DE PUTATIVE ERYTHROCYTE BINDING PROTEIN EBL-1 (FRAGMENT)
GN EBL-1
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID:5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20078864; PubMed-10613703;
RA Peterson D.S., Wellem's T.E.;
RT "EBL-1, a putative erythrocyte binding protein of Plasmodium falciparum, maps within a favored linkage group in two genetic crosses."
RL Mol. Biochem. Parasitol. 105:105-113(2000).
DR EMBL; AF131999; AAD33018.1;
FT NON TER 2647 2647
SQ SEQUENCE 2647 AA; 304551 MW; AE98F88FD754E300 CRC64;

Query Match 3.98; Score 211; DB 5; Length 2647;
Best Local Similarity 19.38; Pred. No. 0.0091;
Matches 245; Conservative 177; Mismatches 441; Indels 406; Gaps 63;

QY 1 MKFSKTYIAASNAVYSLSCAY-----LNHRSEENDNNRVSYVDSQ 46
Db 604 MLKEKYNAGDKICNMLISYADIGDIVRGLDWMDINNRKLEK--FOKITPVGGSN 661
QY 47 SSKSENLTPPOVOSKEGIAQEAIVIKITDQGYVSHGDHYHYGKVP-----YDA 98
Db 662 RKKQNDNERKMKWKKQNLWSSVKKHPIKCKTKRHNNE--KIPQLRLKMGCD 717
QY 99 LPSELLKMDPNYQKADIVNEVKGYYIIVD--GKYVYLLDAHADVNRTRDEINR 155
Db 718 EFCEEM-----GTEVKOLEKICENKNCSEKCKKACSSYEKIKERKNEYMLQSK-RFDS 771
QY 156 OKOEHYKDN--EYKNSNVAARSGRTTN-----DGYF-----NNA----- 191
Db 772 DKLNKKNLNNKFEDESKAYLRSESKOCSNIEFNDETFTFPNKYKKAQMCVENSSEKAL 831
QY 192 ---DI--IEDTGNAYI-----VPH--GGHY-----HYIPKSDLSASE 221
Db 832 KPIKTDVPIETKSKSELSTLDSKNTPNSSGGGNYDROISKRDVHHGPREVYSGE 891
QY 222 LAAAKLAHLAGKNMOPSLSTASDNNNTQSVAKGSYKPKANK-----SENLO 269
Db 892 KEVPRIDAAVK-----TENEFSTNRNDIEGKESKGDHSPVHSKIDKNEEPOQVASELP 947
QY 270 SLKELVDSPS-----AQKYSDDGYVDPDAKIIISTPAGVAIIPGDHGHFIIPYKLS 322
Db 948 KIEKMSSDSIPITHLEAEKGQSSNDNDPAVSGRESKDVNLHSE-----RIK 999
QY 323 ALKEP-----KIAKAVPISGTGTVSTNAKNPNEVYSSIGLSSNPSLSTSEL 370
Db 1000 ENEGVYKTTDDSSKSIEISKIPSDQNNHSDLSQNA--NEDSNOQNKETINPS--TEKNL 1055
QY 371 ---SSASDGY-----IFNPDIYEE-----TATAYVRGDFHFIIPKSNIOG 411
Db 1056 KEIHYKTSDDHSGSIKSEIEPEKLETEESPLTKTRESAAI--GDKNHESYVSADIFQ 1112
QY 412 PTLN-----NSLATPS-----PSLPINPG-TSHEK--HEEDG 441
Db 1113 SEIHSNDRIRIVSESVYVDSGSSMSTESTIRDNKDKFTSEDIAPIINGHEKIGSSADD 1172

```

QY 442 YGPDANRIAEDESGFVMSHGDHN-----HYFCKDLTEBOIKAQKHLEEVAT 490
 DB 1173 KSESDSLIDKSENFENKSSHSIDKQSDNGSTDY---ESLTERSPRG---DLESVSP 1226
 QY 491 SHNGLD-----SLSSHEDYPGN--AKEMDLDKIEEKIA 524
 DB 1227 SSIDMDLKNKSSPPTSSEPHVDSPINSELQASQANADSVQGEKPSKNILRTGDVSEK-- 1284
 QY 525 GIMKOYGVRESIVYKNEKNALITYPGD-----HHNADPIDEKPYGIGHSHSN 573
 DB 1285 -----EKISVSPWAVSYTDEGDKROGISDSSIHHEIDP-----EKNH 1324
 QY 574 YELFKPREBV-----AKKEGNVYTGEE-----LT 598
 DB 1325 YSFSISEGLEBGEIEKEKEDGSLIPSPERKINDGFRREITVDSVERVDNNSDLS 1384
 QY 599 NVVNLKNTFNNQNTLANGORVSPFPELEKLGIM-----LVKLITPD 647
 DB 1385 NVYSEGESEST-----ISSRNGTEGINSELSKSEHTSYDANRDEHNQENLVSSSTQ 1440
 QY 648 GYLVKVSQVGEVGNITANFELDQPLPGQTFKYT-----IASKDYPEVSY 695
 DB 1441 ESERERERKK---ENADSHESELSSISEVGETIRNDAEASENKGEDILQSEGOIYT 1497
 QY 696 DGTFTV-----PSTLAKKASQITIFPFAKD--TILKRVN-QFAPVKGIDALV 741
 DB 1498 ETKKLEPTVNILOPSTPGIEHKESEIDNDKNGEVTHTDAPFYRSESEVAT 1557
 QY 742 RVDEEH-----GNALENNRYKGEIKLIPKLNQGTTRAGNKIPVTMANAYLDNOST 796
 DB 1558 NKPDEHEMTKTPSOIIEK--ELGEI-----DSTINQDN-----DEQGN 1594
 QY 797 YVEVPILEKNOTKPSILPQFRNKAQENSKLDEKVEEPTSEKEVEKLSGTGNS 856
 DB 1595 ST--IFLKNKNEEDGVS--PSTRVMSGSFVSR--NEQIIEEKDKHITDDTTINS 1648
 QY 857 NSTLE---EVPYVDQVQEVAKFAESYGNKLENVLEPMQITIELIPSEVYIKKNADPT 913
 DB 1649 ENGLKGYGMPN-DST--SVVTTTESPLDVEOMLEPIDG-----KKN-EKN-NIT 1694
 QY 914 GAPOGNGENKSENGKSTGTVENOPTENKP---ADSLPEAPNEKVPENSTONGMLN 970
 DB 1695 GEPQESTTIRKOMQPSIN---VNIPELHPVABSKLEBAKE---ISMDDADGTTT 1747
 QY 971 PGANVGS DP 979
 DB 1748 EDITVYEDP 1756
 RESULT 11
 ID 09T274 PRELIMINARY; PRT; 1704 AA.
 JC 09T274
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DE 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE MERZOITE SURFACE PROTEIN 1 PRECURSOR.
 GN MSPI.
 OS Plasmodium falciptarum.
 OC Eukaryota, Alveolata, Apicomplexa, Haemosporidia, Plasmodium.
 OX NCBI_Larid-5833;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-HN2;
 RA Jiang G., Liu R.Z., Daubenberger C.A., Pluschke G.;
 RL Submitted (Apr-1998) to the EMBL/GenBank/DBD databases.
 DR EMBL; AF062348; AAC72885.1;
 DR InterPro; IPRO00561;
 DR Pfam; PF00008; EGF; 1.
 KW Merzoite.
 SQ SEQUENCE 1704 AA; 193762 MW; 385526DDDA56FDID CRC64;

Query Match 3.9%; Score 210; DB 5; Length 1704;
 Best Local Similarity 20.2%; Pred. No. 0.0052;
 Matches 248; Conservative 177; Mismatches 411; Indels 394; Gaps 66;
 QY 1 MFPSKRYIAGSAVYVSLCAVALNQRSQEN---KDNRFVYDGSQSSQKSENLTP 56
 DB 504 MEFNNNF---DKQVYDKIPSAITYNVEKORYNKRFSSNSNSVYV---QKIKKALSTYLE 557
 QY 57 DOVSQKEGIAQIYIKITDQGYTSHGHYHYNGKPYDA--LFSEEL----- 104
 DB 558 D-YSLRKGJ-----SEKDFNHYYLKTGLEADIKKLEIKSENKILE 600
 QY 105 -----LMQPNQLMDADIVN-EVAGGYII-KYDG--KYYVILKDAAHADNRT----- 149
 DB 601 KMFKGLTHSANSLSLESDYKIQVQAVYLLIKKIEDLKLELKLKNAQLKDSIHVRITYP 660
 QY 150 -----KDEINROKO-----EHVKDNEKYSNVA---VARSQGRYTTNDGYVF 188
 DB 661 QMKPEPYIYLAKKEVDKLEETPKYKMDLKEQAVLSSITQPIVAASE---TTEG--- 714
 QY 189 NPADITIEDGNAYIVPHGHHYIPKS-----DISASELAARAHLAGKMAQPSQLSY 241
 DB 715 -----GSHFTLSQSGETEYEETEETEYVGHITYITILPPTOPSP 757
 QY 242 SSTA-----SDNNQSVAKGSTKRPANK-----SENLOSILKELY 276
 DB 758 PREKRVENSIEHKSNDQALTKYLLKDEFLTKSTYCHKYLIVSSSMQKILEY 817
 QY 277 D-SPSAQRSESDGLVDFDAKILSRTPNGVALPHG-----DHYHPIPSK-----L 321
 DB 818 NITPEKNKLS---CDPLDLFNQNNIPAMYSLYDSMNNDIQHLPFLYOKENITYL 873
 QY 322 SALEEK-----IARVPISGTGYST--NAKPNEYVSLGSLSSPSSLTSEKLSMS 374
 DB 874 HLLKEBNHKKILEQKQITGTSSTSPNTVTVAQSAHTHSQOQASMS---TNRQ 930
 QY 375 DGYTFPKQIVETANAYIVRHGDHNYIPKSNQI-GQETLPN--NSLATPSPSLPINF 431
 DB 931 NG-----VAVSGPVAVEESHDPILTVSINDLGIVSLNLGKTYVNP-LTIST- 981
 QY 432 TSEKHEEDGYFDANRIAEDESGFVMSHGDHNHYFKD-----LTEROIKAAQKH 485
 DB 982 TEMEKTYEN-----ILKNDTYF-----NDIKQFVNSKVTIGLETQKNALNDEI 1029
 QY 486 EEWKTS-----HN---GLDSSLSHEDYPGNAKEMD---LDKITEKTAGIMKOYG 531
 DB 1030 KTKLPDLQSLFDLYNRYKTKLDRLEFKKKEKELGQDKQIKKLYLKQLSKINSLNPN 1089
 QY 532 VKRE-SIVYKNEKNAL-----YRPGDHHADPIDEHKPVGT- 567
 DB 1090 VLDNFSVFENKKEAEIAETENTLENTKILNHYKGLVRYNGE---SSPLTLTSEVSIQ 1146
 QY 568 -GSHSNYELFKPEBGVAKKEGNKYVTGEE-----LTNVNLKNTSTFNNQNTLAN 618
 DB 1147 TEDNANLEKPYLSKIDKLNNDLHGKKLSFLSSGHHITELK-EVINKNNT----- 1202
 QY 619 GOKRVSEFPPELEKLGIMLVKL--ITPDGVL-----EYSGVYFGEGGNIN 668
 DB 1203 -----GNSPSENKK--VNEALKSYENFLPEAKVTVVTPPODPTPPLSLVRSVSGSG 1254
 QY 669 FELDQPYLPQ-----TFKTYIASKDYPEVSYDGTFTVPSLAVKMAQSTIYFPFAAGT 723
 DB 1255 STKEQIPTSGLLTELQVAVOSQNTDEED--DSLVLV-----ITGESEDNDE 1302
 QY 724 YL-RVNPQAPVKGTDALVRVDEFGNAYLENRYKGEIKLIPKLNQGTTRAGNKIP 782
 DB 1303 YLDQVYTGALISTYMDNITSGF-----ENRYVYILK-PL-----AGVYRSLKKQJE 1348
 QY 783 ---VTFMANA-----YLD-----NOSTYIYEVPILEKNOTKPS 814
 DB 1349 KHLTFNMLNDLNSRLKRRKYFLDVLSEDLQMFHISNRYIE-DSFKLLNSEQKAT 1407
 QY 815 ILPQFRNK-----AOENSKLDEKY-----EPPKISEKVEKLSGTGNSNSTL 860

```

Db 1408 LKSYTYIESYENDIKFQEGISYEKYLAKYKDDLESIKYIKK---EKEPSSPPT 1464
QY 861 EEPYVDVPOEYAKFA-----ESYGMKLENVLFNM-----891
Db 1465 PRSPATDQKESKLPPLTNIEYLNVLNKKIDYLNLAACINDCAYEADHAKYT 1524
QY 892 -----DGTIELYLPSEGVAKKNNADFTGEAPOGNGENKPSNGK-VSTGTVENOPT- 941
Db 1525 KLSDDKAIDKIDLF-----KNTNDFEALIKKLINDTKDKMGLSTGLVGNKFPPT 1576
QY 942 -----ENKPADSLPEAPNE--KPYKPPNS 963
Db 1577 IISKLEGFQDMNTISOHCYKQCPNS 1606

RESULT 12
ID 025922 PRELIMINARY; PRT: 1720 AA.
AC 025922;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PRECURSOR OF THE MAJOR MEROZOITE SURFACE ANTIGENS.
DE Plasmodium falciparum (isolate NF54).
DE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
[1] NCBI_TaxID=5843;

RC SEQUENCE OF 1069-1720 FROM N.A.
RA STRAIN-NF54;
RL Tolle R., Bujard H., Cooper J.A.;
RL Exp. Parasitol. 0:0-0(1995).
[2]
RC SEQUENCE FROM N.A.
RA STRAIN-NF54;
RL Tolle R.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
[3]
RC SEQUENCE FROM N.A.
RA STRAIN-NF54;
RX MEDLINE=96123395; PubMed=8577332;
RA Pan W., Tolle R., Bujard H.;
RA "A direct and rapid sequencing strategy for the Plasmodium falciparum
antigen gene gp190/MSA1."
Mol. Biochem. Parasitol. 73:241-244(1995).
RL EMBL: 235327; CA84556.1;
CR InterPro: IPR000561;
CR Pfam: PF00008; EGF; 1.
CR Signal: Metazoa;
CR SIGNAL 1 19
CR SEQUENCE 1720 AA; 195725 MW; 717B2FD1E637A8A3 CRC64;

Query Match 3.8%; Score 205.5; DB 5; Length 1720;
at Local Similarity 20.1%; Pred. No. 0.0094;
atches 248; Conservative 171; Mismatches 413; Indels 403; Gaps 65;

```

```

QY 189 NPADIEDTGNAYIVPHGHYIPKS-----DLASSELAAAHLAGKNMPSQISTY 241
Db 730 -----GSHHTLSQSGETEVEETEELVEVGHITVITLPTLPSP 772
QY 242 SSTA-----SDNTQSYAKGSKSPAK-----SENQSLKELX 276
Db 773 PREVEVENSIEHKNSDQALTKVYLLKIDELTKSYICHKYLIVSNSMDOKLLEY 832
QY 277 D-SPSQRYSSEGLVFPKATISRTPNGVALPHG-----DHYHIFYSK-----L 321
Db 833 NITPEENELKS-----CPDLLEFNQNNIPAMYSIDYSDMNDLHLEFELYOKMITYL 888
QY 322 SALEEK-----IARVPISGTSTVST-NAKPNEVSSLGSLSSNPSLTTSKELSSAS 374
Db 889 HKLKEENHKILLEEOKOITGSTSTSPGNTVTNAGATHSNSQOOSMASS--TWQ 945
QY 375 GYINPNPDYEEATATVVRGDHFHYIPKSNQI-GQPTLPN--NSLATSPSPILPG 431
Db 946 NG-----VAVSSGPAAVESHDPLTVLSTNSDLKGLVSLINLGNKTKVPNP-LTIST- 996
QY 432 TSEKHEEDGYCFDANRLIAEDSGFVASHGDHNYFPKID-----LTEDQIKAAOKHL 485
Db 997 TEMEFYEN-----ILKNNDYF-----NDIKQFVKSNSKVITGLTEQKNALNDEI 1044
QY 486 EEVKTS-----HN-----GUDSLSHEDYFGNAKEMKD--LDKTEKIAGIMKOYG 531
Db 1045 KKLKDTQLSFDLYKRYKILDRLENKKELQDQKQIKLTLLEQESKLINSINPDN 1104
QY 532 VKRE-SIYVNEKNALII-----YPHGDHNDPIDENKPGI- 567
Db 1105 VLQNSVFNKKKEALIEETENTLTKILKHYGLVKNYNGE--SSPLTTSSEVSIQ 1161
QY 568 -GHSNRYELFKPEEGVAKKGNKYTGEE-----LTWVNLKSTENNONFTLAN 618
Db 1162 TEDNANLEKRYVLSKIDKLNLDNLHCKKLSFLSSGLHHTITLK-EVKNKNYT--- 1217
QY 619 GOKRVYSFPEPELEKILGNLTKL--ITPDGKYLEVSGRVGEGVGNIANEELDQY 675
Db 1218 -----GNSPSENNKK--VNEALKSENFLEAKVTVTVP-----PGPD 1254
QY 676 LPQGFRTYI-----ASADYPEVSYDGFYPTSLAYMAAQITFYPPHADDTYLRVNPQ 730
Db 1255 VTPSPLSVAVSGSSGSTKEETQIPPSG-----SLTTELQOVVOLQNTDEDDSLVYLP 1308
QY 731 FAVPRGTD-----ALVRVDEFGNAVLENNKYVGEIKLPRLNGGTTRTAGNK 780
Db 1309 FGESENDDEYLDQVYTGALISVTMDNLSG--FENEYDVIYIK-PL--AGVTRSLKKQ 1361
QY 781 IP-----VTFMANA-----YLD-----NOSTYIYEVPLLEKENDTDK 812
Db 1362 IEKNIFPNNLNDLINSRLKRRKRYFDVLESDLMQFHNISNEITIE-DSFKLLNSQOK 1420
QY 813 PSILPQFRNRK-----AOENSKLDEK-----EERKSEYKVEKKLSETGNST 855
Db 1421 NTLKRSYTYIESYENDIKFQEGISYEKYLAKYKDDLESIKYIKVEKKFPSSPPT 1480
QY 856 NSNTLEEVPTDPOEYAKFA-----ESYGMKLENVLFNM-----891
Db 1481 PPS-----PAKTDAQKESKLPPLTNIEYLNVLNKKIDYLNLAACINDCAYEADH 1535
QY 892 -----DGTIELYLPSEGVAKKNNADFTGEAPOGNGENKPSNGK-VSTGTVE 937
Db 1536 HVKITKLSDKAIDKIDLF-----KNYDEALIKKLINDTKDKMGLSTGLVQ 1587
QY 938 NQPT-----ENKPADSLPEAPNE--KPYKPPNS 963
Db 1588 NFPTIISKLEGFQDMNTISOHCYKQCPNS 1622

RESULT 13
QYLBG3 PRELIMINARY; PRT: 1653 AA.
AC QYLBG3;

```

01-OCT-2000 (Tremblrel. 15, last sequence update)
 01-MAR-2001 (Tremblrel. 16, last annotation update)
 DE PAA.
 GN Streptococcus criceti.
 OC Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 NCBI_TaxID=1333;
 RN SEQUENCE FROM N.A.
 RP STRAIN=849;
 RA Tamura H., Kato H.;
 Cell surface antigen I/I - Streptococcus cricetus.
 Submitted (Apr-2000) to the EMBL/GenBank/DBD databases.
 EMBL: AB042239; BAA95000.1;
 DR InterPro: IPR001899;
 DR InterPro: IPR002965;
 DR Pfam: PF00746; Gram_pos_anchor.1.
 DR PRINTS: PRO1217; PRICHEXTENR.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN.1.
 SO SEQUENCE 1653 AA; 18062 MW; 93F2B215BF598DE CRC64;

Query Match 3 7%; Score 198.5; DB 2; Length 1653;
 Best Local Similarity 18.6%; Pred. No. 0.021;
 Matches 242; Conservative 166; Mismatches 435; Indels 459; Gaps 63;

2 KFKKTYIA--GSAVYSLICAVYALNQHRSQENKDNRYSYD-----GSQSSQKSENT. 54
 13 KIKSLGALGTAIYVSA-----GQRLADMTSTSAVDTAIVGTETGNGPATNL. 66
 55 ---TPQVSOKEGLOAEQIYKITTDOGYTSHGDHYHYNGKVPYDALSE----- 102
 67 PERQADSSQAEKSOAQA-----EOKTGMPPVAVATTELDIAVYSSAA 108
 103 -----ELKKDPRYQKADIVN-EYKGGY-----IITKDGKYYVYLKQAAHDVY 147
 109 EAGVYVQSDTVKNGVYSGQDEKSGEIKADYSKQAEYIKTTETV-----KADVA 161
 148 RTDEINRQKQEHVKNENKYNVAVASQGRYTTNDGYFNPADITIEDGNAYIYBGG 207
 162 ANQSENRINQENAAKQAYEODLANKAKEVEITNK-----NQ----- 201
 208 HYHYIPKSDISA-----SELAAR-----AHLAKNMOPQSUSYSTASDNTQSY 253
 202 -----AKADYEAKIAGYOKELAAVQOANSNDQAAYAAKADAYEKELARVQKA---NADAK 253
 254 AKGSTSRPAKSENQ-SLILKEIYDPSAQRYSQGLVDPKAIISRTNGVAIPIHGDH 312
 254 AAYEQVVAANTATNEQIKANMAIIOQRMAQADYEAKIAGYOKDLAAVQAGNAHEADY 313
 313 YHPIPSKLSALEEKIARVPISGTST-----VSTNAKNEVYSSLSGSSNPSSITTS 367
 314 -----QAKTAYQDELARVQANAAKTAVEQVVAANTANAEQIKANNTAIOQRMAQKA 368
 368 -----KELSSASDGYIFNPDI-VEETA-----TAYIVRRGDHP 400
 369 DYEAKLAYQNDLAAVQAGNAHEADYQAKKTAYEDELARVQANAAKAAKAY----- 420
 401 HTYIPKSNQIQPLPNNISLATPSPSLPIFGTSHKEHEEGGYFDPANRIIAEESGFWMS 460
 421 -----EQAANAANTAKKEEIKANEA--IKORNAQKADYEAKIAGYOKDLAAVQAGNA 473
 461 HGDNHNYEFKDLTEROIKAKHLEEVKTSHNGL--DSLSSHQ----- 503
 474 EAD--YQAKKTAYDEELARVQKANADAKATYQAVKDNQAKNAETIAEAAAIROKRAAA 530
 504 --DYPSN-AEMKDL-----DKRT--EEKIAGIKQYGVNR-----ESTIV 539
 531 KADYEAKIAGYOKDLAAVQAGNAHEADYQAKKTAYETELARVQKANADAKAAVDAQVKD 590
 540 NKEKNALITYHG----- 551

DB 591 NQAKNNEIAENNAIQRNAAKADYEAKIAGYOKDLQYOKDFADYQSKLXAYEDAQA 650
 QY 552 -----DHHHAD-----FIDEKPPVIGHSNHYLFEPBESVAK--EGNKVYGEEL 597
 DB 651 YKAAIALEKHKNDGDLTSPESQPF-----IFQENKATLETISGQVYNADEF 700
 QY 598 ---TNYVNLKNSI-FNNQFTLA-----NGQKRVSPFPELEKKGIMLVYLIPDCK 649
 DB 701 YAEADKLIDYKADLTLENTERRRAYEHPKGGIDSDADWPK-----RVFLPKGR 749
 QY 650 -VLEKSGRVFEGGVGNINAFELDQPYLPQGEFYKYLASRD-----YPEVSYDGTFTV 701
 DB 750 PIVARYS-----NLKNSLNGKIKIASAEPIYTL--KDTSLNDKIPAMLYD----- 794
 QY 702 PISLAYKMASQITFYFPNAGDYLRNPPQAVYK-----TDLVAVFDFEFGNALLENN 756
 DB 795 PL-----QTIYFEFPGSOSIDVDIKFYEDGKNKIDATGALIDFSSLSNRHGMPQYN 846
 QY 757 YKGEIKLPIPKLNGSTTRAGNKIPVTFMANAYLDSQTYIYEVPILEKNO----- 809
 DB 847 -GINTLE-----KVNHYT-----GTALIN--GSSVKDNGDGLYAATNTKANSRESNE 895
 QY 810 ---TDRP-----SIL-----POFKRNKAQENSK-----IDKEYEPTSEKYEKELSET 851
 DB 896 WDTNPSVMYGAIVGKQGPDIKFTLSAENAGIWFAPDSKI-----KAKDI 943
 QY 852 GNSTNSLTLEVPYVDPQEKVAKFAESYGMKLENLFNMDGITELIYDSGEYIKKNMAD 911
 DB 944 PNKPMES--EPTPPVAPQ-----PLE-----PYEVEK-----D 971
 QY 912 FTGEAPQGENKSPENKGYSTGVENOPTE-----NKPADSLPEAPN--EKPVPENST 964
 DB 972 LEFVPEVNPDPPTPTKPPDQPEPNKPDEPNYEYKESSEAPPIENPDPEPTPTKTP 1031
 QY 965 DNGMLN-----PEGNVGSDPMLDPALAEAVDPVQDEKL----- 998
 DB 1032 DQPEPNKPDEPNYEYQGPAPAVEPKYD--EPTPPVDPYEKIKIDPVVAVYRHYHK 1089
 QY 999 -----EKTASYIGLDSVIFNMDGITELR-----LPBG 1027
 DB 1090 LAIQPGVTAKIKNSDMLDIDKTLVAKOSTVFKFOLLADLPAG 1131

RESULT 14
 Q9LC00 PRELIMINARY; PRT; 1698 AA.
 AC Q9LC00;
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
 DE HYPOTHETICAL 181.9 KDA PROTEIN.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 NCBI_TaxID=1280;
 RN SEQUENCE FROM N.A.
 RP STRAIN=NCIC10442;
 RA Ito T., Hiratsuki K.;
 RT Comparison of three types of Srmec.
 RL Submitted (Oct-1999) to the EMBL/GenBank/DBD databases.
 DR EMBL: AB033763; BAA86640.2;
 KW Hypothetical protein.
 SO SEQUENCE 1698 AA; 181885 MW; E209D67161F9D84B CRC64;

Query Match 3 7%; Score 198; DB 2; Length 1698;
 Best Local Similarity 18.9%; Pred. No. 0.024;
 Matches 228; Conservative 144; Mismatches 465; Indels 372; Gaps 50;

2 KFKKTYIAGSA-VIYSLIC-----AYALNQHRSQENKDNRYSYDGS-----QS 47

```

Db 18 KYSIRFVGTASILVGAFLIGVANDQAAEANNNTTOKODSSDKVKNVOTIESS 77
2Y 48 SOKSENLTPOVSQKSGIQAEQIVIKITDOGYVTSKDHVHYNGKVPYDALFSEELAK 107
2b 78 ANSNEDIEQVD-----VTKDTE-----QASIEKANNT 108
2Y 108 DRYOLAKADIVNEVKGITIVKDGKYYLLDADAHAD-----NVRKDEINROK 157
2b 109 EQASTEERKADTEQATTEAPRAEGTDKETEAPRAEETDKATEAPTEETDKATEE 168
2Y 158 QEBVRDNEVNSVAVASQGRRTNDGVFNPAIDIEPTGNAYIYBGGHYHYPKSD 217
2b 169 AKRAETDIAETAPTEETDKATEE-----APAEETSKAAT-----EBAKAE 215
2Y 218 S-----ASELAAKAHILAKKNMOPOL 239
2b 216 TSKAATEAPKAETEKATTEAPKTEETDKETEAPAEETSKAATEKAKRAEETNV 275
2Y 240 SYSSSTADNNTQSVAKGSKPANKSENLOSLK-----ELYDSPAORISESDG 290
2b 276 ETEAPAEETKAAEET--PAVEDTNAKSNMOPSETERTOVVDYAKOLYKSEVT 333
2Y 291 VEDPAKIISTRPAGVAPRGGDHYHFIPYSKLSALE-EKIAVAVPISGSGVSTNAKPE 349
2b 334 EAKAEIEKVLPRDI-----SNLSNEEKIKALSEVLAKETAN--KENAOPRA 378
2Y 350 VVSLGSLSSNSLTSKELSSASDGYIFNPKDIYEETATAYIVRGGHFIYIPKSNQI 409
2b 379 TRRSVSNARTNVNSAVALRAAA-----ODVTKKGQNFHAGDIIHKYKEEP 431
2Y 410 GQPTL-----PNNSLATP-----SPSLPINTGSHKEEGCYGPDANR 448
2b 432 NGCTLAFTNFNPTGTALEYENDKIDFNEDFITVAVANNNGNTGAGWGWMFQ 491
2Y 449 IIAED---ESGVMSGHDNHYFFKDLTEEOIKAAOKHLEKTSNGLDLSSEDOY 505
2b 492 GNGOFLNGGILRKGANASGFKIDTAYNNVNGVKDLDDKT--NLSQIGAKVGY 549
2Y 506 PGNNAEMKDLKIEKIKAGIMKOYGVKRESIVYKNE--NAILIPRGGHHAIDPEHK 563
2b 550 GTFVKNAGD-----GYTNQVG--QNALNTRKDKPVNKIIVADNTNHL 590
2Y 564 PVGIGHS-----NYELKPEBGVAKKGNVYGEELTVNNVLKSNFTNNQNT 615
2b 591 ---GOFHQRLNDVYLAND-----AATSTIYAKTKATTDIDGIDKSKYNTL 639
2Y 616 LANGORNVFS---FPPELEKLGIMLKYLTTPDGKYLEKVSAGVGGVGNIANFELD 672
2b 640 ITSSHONKNSGIMKRTNLES-----VTITPOADLIDV--EYTKOPIRKTIRED 690
2Y 673 QPYLPGQFFKTYIASKDYEV---SYDG--TFVPTSILAYKMASQITFFPHAGDTYLR 726
2b 691 PLEBGS-----PDVYVKGEGDEKTTTPT--KVDPDT-----GDVER 728
2Y 727 VAPORAVPRG--TDALVYR-----DEFHGNAYLENNKYVGEIKPIPLKNGTIRT 776
2b 729 GEPTVEVKNPVDLIVHAPAEVPOGHKDEDPN-----LPI-----DGEVY 771
2Y 777 AGNKIPVTEMANAYLDNSTYIYEPILEK-----ENOTDKPSILPOKRRKAQEN 827
2b 772 PG-----KPGIKNPEITGVYTPPVDDVTKHOPKAGEPEVYKEELPEEKRRFENDL 822
2Y 828 SKLDEKVEPTSEKVEKEKLESTGNSTNSNLEEV---PTVDPVQEKVAKFAESYMK 883
2b 823 KPGEEKV---TOEGQTEKTTTPTTINPLLGKVGEGEPTTEKTPKPVDEIQFGEE 878
2Y 884 LE-----NYLFNDGDIEXLP-----SGEVAKKMAADTGAPGNC----- 920
2b 879 VPQGHKDEFPNLPIDGEE--VPKPGIKNPEGEVYTPPVDDVTKHOPKAGEPEVYKE 936
2Y 921 -----GENKPSENGKYSTGY-----ENQPT--ENKPA 946
2b 937 ELPPEKKEFPNDLKGEEKYTOEGQTEKTTTPTTINPLLGKVGEGEPTTEKTPKPV 996

```

```

QY 947 DSLPAPNEK-----PYKPNSTDN-----GMNPE-GNVGSDPMLDPALEAPA 990
Db 997 DEITQGGEEVPOGHKDEFPNLPIDGTEEVYKPGIKNPEGEVYTPPVDDVTKHOPKA 1056
QY 991 VDPVEKLE 999
Db 1057 GEPEVYKEE 1065

RESULT 15
Q9RFJ4 PRELIMINARY; PRT; 881 AA.
AC Q9RFJ4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE SERUM OPACITY FACTOR PRECURSOR (FRAGMENT).
GN S0F1207.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1114;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S1457;
RA Beall B.;
RT "Predictions of M serotype, anti-opacity factor type, and highly
related strain sets based upon emm and sof sequence types."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF191035; AAF23117.1;
DR InterPro; IPR002035;
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW Signal.
FT NON_TER 1 1
FT SIGNAL <1 12 POTENTIAL.
FT NON_TER 881 881
SQ SEQUENCE 881 AA; 97132 MW; 64B42CFB06C564D1 CRC64;

Query Match 3.6%; Score 196.5; DB 2; Length 881;
Best Local Similarity 20.8%; Pred. No. 0.01;
Matches 194; Conservative 123; Mismatches 371; Indels 245; Gaps 40;

QY 172 AVASQGRRTNDGVFNPA--DIEDFGNAYIVPHGGHYHYPKSDLSASELIAAKAHL 229
2b 24 AGATTNGATTITGTAAGTAAGSAGSESSDASVPSGS-----QSSSAPA----- 70
QY 230 AGKNMOPSLSYSTASDNNTQSVAKGSTSKPANKSENLOSLKELVD--SPAQRYSSED 288
2b 71 ---SPOQAQVTVATTSYSTSSSDGKSTKATSTSSASTSSSEDKAKRAASTKSS 127
QY 289 GLVYDPAKIISTRPAGVAPRGGDHYHFIPYSKLSALEKIAMVPISGSGVSTNAKPN 348
2b 128 TTVASPS-----NG-----SNQGSTETEDPM 150
QY 349 EVVSSLSGLSNPSSILTSKELSSASDGYIFNPKDIYEETATAYIVRGGHFIYIPKSNQ 406
2b 151 EVEQYKDKETELKADGSKLNSSDKILRAND--GEQDIDIDIRE-----YKINS 202
QY 409 IGQPTLPNNSLATPSPILPINCYS-----HEKHEDGYGFANRI---IADESGFV 458
2b 203 DG--TIDVTVVYTRK--EIDEGADYVALLDVSKMSSEDPNNAKKIKLVTLTKST 257
QY 459 MSHDHN-----HYFFK-----DLTEQIKAAQKHL--EYVTSN--GLDSLSSEED 504
2b 258 DSGPNHNAKNSVRLATYRYKAYNEPIELAEVNDAKLEWVMOAKKMDWGV-----D 309
QY 505 YPGNAKEMKDLKIEKIKAGIMKOYGVKRESIVYNNKANAIIYPRGDHHAIDPEHKP 564
2b 310 LGQIHAKRELFEKKEKNSG-----KQOHVLFQO----- 340

```


Thu Sep 27 09:15:59 2001

us-09-471-255-2.rspt

Page 13

QY 565 VGJHSNSNTELFREPEGVAKKBNKTKYGEELJNYYNLLKSTFNNOPTLLANGKRS 624

Dd 341 ---GEAFSEYDINDKONTYK---NRI-TGKVTTSNPEFLMIDFHNHTN-----OKADM 387

QY 625 FSPPELEKKGIIMLTKLITPDG-----VYLEKVSQKVGSEVGNJANIELDOPIPCG 679

Dd 388 LDDJAKYIKKLKDGVSYSVSTADSVLSALMLNGFSFPLGSMTEYITLKEYOSDLEEK 447

QY 680 TEKYTISKADYPEVSYDGTFTVPTSLAYKASOTIEFYFHAGDTYLRVNQFAVPKGTDA 739

Dd 448 YEDYT---KRVGEGYHSHS-----NKRSEOM-PF---ESQKATLOTALPKFKDS 493

QY 740 LVRFDEPHGNATYENNYKKGELKLPYKLNQGTFRAGKIVYTPAANA--YLDNOSTY 797

Dd 494 -YSWFAOILBGLGISNOEQAKOAVYIKILN-----TLYKROHYHNHLSA 540

QY 798 IVEVPILEKENOT---DKPSILPOFRNKAOENSKLDEVEEKPTESEYK-----E 846

Dd 541 IAEKMAOEETITYSVDTYELKAPKRVROQAVRKDOTEENKKNKEEERNAKFPDYLK 600

QY 847 KLESTGNST-----NSTLEEVPVADVOEKVAFSAESYGMKLENYFLNDDGTIEL- 897

Dd 601 KMSDEGNFLSVBERDKKXOTYLTFLTIKDEFTDKVAKYQDESKKKYTSL--ISGINSNY 658

QY 898 -----YLPSEVYIKKMMAD---FTGEAPQONG-----ENKPSENGK 930

Dd 659 HTSARSGFSSKESILWTJISMEQLKAFEDGKPLTLTYKLVNDKLLKALEDKKKKKAK 718

QY 931 VSTGTV-ENOPTKPKPADSLPEAPNEAP-----VAPENSTDGMALNPEG 973

Dd 719 HSTSTENENSVTEKTIINSITTYOINOKKGDGKSLGDVNLTYSKFYKVPVPOIDQVIEPO- 777

QY 974 NVGSDPMLDPALEAPVADPOVEKLEKFTASYG 1006

Dd 778 -APKLPDLPPTENGVPALDTTEESITYLPLEHG 809

Search completed: September 26, 2001, 22:33:45
Job time: 264 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 22:27:16 : Search time 46.65 Seconds
(without alignments)
1696.578 Million cell updates/sec

File: US-09-471-255-2

fect score: 5406
Jence: 1 MKRSKRTIAGSAVIVSLSL.....IELRLPGEVYIKKMLSDPIA 1039

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query | Match | Length | DB | ID | Description |
|------------|-------|-------|-------|--------|--------|----|---------------------|
| 1 | 921.5 | 17.0 | 822 | 2 | T46758 | | hypothetical 92.4K |
| 2 | 230 | 4.3 | 2485 | 1 | H71621 | | serine/threonine-s |
| 3 | 226.5 | 4.2 | 1701 | 2 | A54498 | | major merizolite su |
| 4 | 218 | 4.0 | 1873 | 2 | T30944 | | surface protease pr |
| 5 | 215.5 | 4.0 | 1701 | 2 | A26868 | | major merizolite su |
| 6 | 214 | 4.0 | 1939 | 2 | T18372 | | repeat organellar |
| 7 | 205.5 | 3.8 | 1185 | 2 | A42404 | | collagen adhesin - |
| 8 | 205.5 | 3.8 | 1726 | 1 | SAOQM | | major merizolite su |
| 9 | 203.5 | 3.8 | 1849 | 2 | C41859 | | Iga-specific metal |
| 10 | 202.5 | 3.7 | 1726 | 2 | A45948 | | major merizolite su |
| 11 | 196.5 | 3.6 | 1271 | 2 | A45555 | | glutamate rich pro |
| 12 | 196 | 3.6 | 5005 | 2 | F82884 | | hypothetical prote |
| 13 | 193 | 3.6 | 3724 | 2 | T18427 | | hypothetical prote |
| 14 | 190.5 | 3.5 | 4549 | 2 | T20771 | | hypothetical prote |
| 15 | 190.5 | 3.5 | 4667 | 2 | T20774 | | hypothetical prote |
| 16 | 187 | 3.5 | 1080 | 2 | T43164 | | Iacta protein - Ll |
| 17 | 187 | 3.5 | 4688 | 2 | F82885 | | hypothetical prote |
| 18 | 186.5 | 3.4 | 1570 | 2 | T18272 | | 1-phosphatidylinos |
| 19 | 186 | 3.4 | 1331 | 2 | T28681 | | rib protein - Stre |
| 20 | 185.5 | 3.4 | 1071 | 2 | E85343 | | hypothetical prote |
| 21 | 185 | 3.4 | 2195 | 2 | S61103 | | SEC16 protein - ye |
| 22 | 184.5 | 3.4 | 1466 | 2 | A36426 | | SP22 protein - ye |
| 23 | 184 | 3.4 | 3381 | 2 | T43389 | | verecan precursor |
| 24 | 182.5 | 3.4 | 1023 | 2 | T48997 | | epsin-like protein |
| 25 | 181 | 3.3 | 1879 | 2 | UW0059 | | ncpud protein - mo |
| 26 | 181 | 3.3 | 1983 | 2 | G86643 | | hypothetical prote |
| 27 | 180.5 | 3.3 | 5170 | 2 | T15348 | | hypothetical prote |
| 28 | 179.5 | 3.3 | 1315 | 2 | T28679 | | fibrinogen-binding |
| 29 | 178 | 3.3 | 1436 | 2 | S57238 | | forced protein 5.4 |

| | | | | | | | |
|----|-------|-----|------|---|--------|--|----------------------|
| 30 | 178 | 3.3 | 1449 | 2 | S57237 | | forced protein 5.6 |
| 31 | 177 | 3.3 | 1702 | 2 | A41859 | | Iga-specific metal |
| 32 | 176.5 | 3.3 | 1139 | 1 | E64234 | | Cytadherence-acces |
| 33 | 176.5 | 3.3 | 1532 | 2 | A26036 | | Iga-specific metal |
| 34 | 176.5 | 3.3 | 1566 | 2 | A43607 | | cell surface antiy |
| 35 | 176.5 | 3.3 | 1658 | 2 | S55101 | | hypothetical prote |
| 36 | 176 | 3.3 | 3147 | 2 | T18674 | | hypothetical prote |
| 37 | 175.5 | 3.2 | 2401 | 2 | T28676 | | rhoptly protein - |
| 38 | 175.5 | 3.2 | 2748 | 2 | S57976 | | nuclear migration |
| 39 | 174.5 | 3.2 | 1577 | 2 | A35140 | | hemolysin A precu |
| 40 | 174.5 | 3.2 | 1807 | 2 | S03124 | | vitellinogenin A2 pr |
| 41 | 174.5 | 3.2 | 2294 | 2 | I67630 | | protein tyrosine p |
| 42 | 174.5 | 3.2 | 2466 | 2 | I67629 | | protein tyrosine p |
| 43 | 174.5 | 3.2 | 2490 | 1 | A54971 | | protein-tyrosine p |
| 44 | 174 | 3.2 | 1631 | 1 | SAZOK1 | | major merizolite su |
| 45 | 173.5 | 3.2 | 2722 | 2 | T20532 | | hypothetical prote |

ALIGNMENTS

RESULT 1
T46758
hypothetical 92.4K protein - Streptococcus agalactiae
C1Species: Streptococcus agalactiae
C1Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
C1Accession: T46758
R1Spellerberg, B.; Rozdzinski, E.; Martin, S.; Weber-Heymann, J.; Schmitzler, N.; L
Infected. Immun. 67, 871-878, 1999
A1Title: Lmb, a protein with similarities to the Lrat adhesin family, mediates attach
A1Reference number: 224091, PMID:99115568
A1Accession: T46758
A1Status: preliminary; translated from GB/EMBL/DBJ
A1Molecule type: DNA
A1Residues: 1-822 <SP>
A1Cross-references: EMBL:AF062533; NID:94249632; PIDN:AAD13797.1; PID:94249624
A1Experimental source: strain R268
C1Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

| | | | | |
|-----------------------|------------------|---|------------|------------|
| Query Match | 17.0% | Score 921.5 | DB 2 | Length 822 |
| Best Local Similarity | 26.9% | Pred. No. 5e-38 | | |
| Matches 267 | Conservative 152 | Mismatches 305 | Indels 267 | Gaps 31 |
| QY | 1 | MKRSKRTIAGSAVIVSLSCAALNQHRSQENKNNRVSIVVGSQSSQKS--ENLTPDQ | 58 | |
| DB | 1 | MKRTYIGSVAAIILATHIGSYOLGKHHGLATKQNLAVIDSGKVAAPRTKNTMDQ | 60 | |
| QY | 59 | VSOKEGLOAEQIVIKITDQGYVTSRGDHYHYNGKVPYDALFSEELMKDPNTQLDADI | 118 | |
| DB | 61 | ISEEDISAQIVKITTDOGYVTSRGDHYHYNGKVPYDALFSEELMKDPNTQLDADI | 120 | |
| QY | 119 | VNEVKGKIIKVDGKYVYLKDAHADNVTKDEINRQKQEHYD--NEKVSVA | 172 | |
| DB | 121 | INILIGLYIKVNGNYVYLKPSGRKRNITKQIAEQVAKGTEKKEKLAQVAVLSKE | 180 | |
| QY | 173 | -----VARSQRTTNDGVENFPADIIETGNAYIVPHGCHHYIPKSLSELAAR | 226 | |
| DB | 181 | EVAANVEARQKGRYTTDGTIFSPYDIDLDGAVYVPHGNHNYIPKDLSPSELAARQ | 240 | |
| QY | 227 | AHLA---GKNMPSQSLSYSTASDNTQSVAKSGTSKPAN----- | 263 | |
| DB | 241 | AYMSQKQKGRARS--DYRTPAPGRKAPIPVYTNPGGCHQPDNGVYHAPPRNDAS | 298 | |
| QY | 264 | -----KSENIGSLLELYDPSAQRYSSEGLVFPDAKTIISRTPNCAVIPHGDHY | 314 | |
| DB | 299 | ONKHQDEPKGTFKFKELDQHLRLDLKYRHYVEDGILFEPTQVYKSNAGGYVPHGDHY | 358 | |
| QY | 315 | FIVYSLSALAEKIAANVPISGSGVSTAKNEVYSLSGSSNPSSLTSSKELS--- | 371 | |
| DB | 359 | IIPRSLSPLEMLADRYLAGOTDNDSS-----GSDHSKPS-----DKETVTHF | 402 | |
| QY | 372 | -----SASDGYIFNPKDIVEETATAYIVRHGDHRYIPKSNQIQOPTL | 414 | |

```

3b 403 LGHRKAVGKGLDGRPYDSPAAYVSKESHSVDKSGYAKGDFPHYI-GGGELEQVEL 461
3y 415 PW--NSLAPSPSLINPGTSHEKEEDGYFD-----ANRIAEDESGPVMSGHGHNHYE 468
3b 462 DEVAMWVAKGQADELVALLDOEGKERPL-FDIYKVSRTKDKDQVGLMKDQKDFY 520
3y 469 FKRDLEIKAAQKHELEVKTSANGLDLSLSHEDYFGNAKEMKDLKIEKTAGIMK 528
3b 521 ARYQDLQIAFAEDELMLKDKKHRYDI-----VDTGIEPLAAYVS 563
3y 529 QYGVRESIVYVKEKNALITYPHGDHHPD---IDEKPVGISHSHNTELEKPE--EGV 583
3b 564 SLPMAGNAYDTGSSFYI-PAIDIHIVVPSWMLRNQIATF---KYVQHPVPRDY 617
3y 584 AKKEENKTYTGELNLYNVLKNSFFNNQNTPLANGOKRVSPFPELEKTKIGIMLVKL 643
3b 618 WSKPQHE-ESSGVITNVPDLKRAQPMQI-----IHSAEVQKALAEQ--RF 663
3y 644 ITPDGVLEKVSQVGVGNIANFELDQPYLPQOTERYTASQDEVSQGTFTVPT 703
3b 664 AAPDGYFD-----PRDYLAKEF-----VMDGGSFISPR 693
3y 704 SLAYKMAQSTIYPPHAGDTYLKVPQAVPKGTALVRFDEFGNAYLENNYVGEIK 763
3b 694 A----- 694
3y 764 LPPLNCGTETKATNKIPVTFMANAYLDNOSTYVEVILEKEN-----QTDKSLIPQ 818
3b 695 -----DSSSLRTNKDL-----SQAEQQAQDELAKKAGNATDIDKEEKQ 738
3y 819 EKRNKAQNSKIDEEVEEPTSEKYEKELSETGNSNSTLEEVPTD---PVQEKVA 874
3b 739 -ADRSNEN-----QQPSEAKSEKE-----SDDFIDSLPDVGLDRATEJEDIN 780
3y 875 KFAEYQKLENVLEFMDGTIELIYPSGEVI 905
3b 781 QLAQKANIDPKLIFQPEG-VQFYKNGELV 810

RESULT 2
471621
serine/threonine-specific protein kinase (EC 2.7.1.-) PFB0150c - malaria parasite (Plas-
)Species: Plasmodium falciparum
)Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
)Accession: H71621
)Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
)Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
)Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
)Reference number: A71600; MUID:99021743
)Accession: H71621
)Status: preliminary; nucleic acid sequence not shown; translation not shown
)Molecule type: DNA
)Residues: 1-2485 <GAR>
)os-references: GB:AE001376; GB:AE001362; NID:93845108; PIDN:AC71820.1; PID:9384510
)Experimental source: clone 307
)Gene: PFB0150c
)Superfamily: malaria parasite serine/threonine-specific protein kinase PFB0150c; prote
)Keywords: phosphotransferase
)2087-2352/Domain: protein kinase homology <KIN>

Query Match 4.38; Score 230; DB 1; Length 2485;
Best Local Similarity 18.08; Pred. No. 0.0042;
Matches 209; Conservative 154; Mismatches 397; Indels 400; Gaps 47;

3y 12 SAVVSLICAYALNQHSGENKDNRRVSYVDGSSQSKSNLTPDVOYSGKQIGIAQIV 71
3b 553 SISLNNVFFYKKRKSSNNNNNNNNIS--SSSSSSKKNHVI--INRK----- 598
3y 72 IKITDQGVTVSHGDHYHYNGKVPYDALFSEELM----- 106

```

```

Db 599 -----ISSYNIHYKERKDSFKENPLFFKFKILPSKKDQVFNERRQKDEKSNHIK 650
3y 107 -----KDPNYQLKADIVNPKGYYIIKVDGKYVYLKQAAHND 146
3b 651 CVSSFNNTSDISSHSYNNKEPFPALNNISRIHPKNNIITYSGKSFNVQDEKTVL 710
3y 147 VRTKDEINROKOEYKDNKYNNAVARSGRYTNDGYFNPADI-----IEDTG 198
3b 711 LKKKEIN--DKNFFSCLINHTT-----YLGQVKNLNMGLRDSIYKIDEN 761
3y 199 NAYIVPHGHYHYIPKDSLASELAFAHLAKGNMOPQSLSSTASDNTQYAKGST 258
3b 762 NMLKECYGNNDNNKKKKKKKLSFS-CDIINDNITP-----YESDKKNSNNKSMDI 816
3y 259 SKPANKSNLSLKELVDS--PSAQRSESDGLVFDPAKIIISFPNVAALPHGDHYFI 316
3b 817 FNYVAKSNLNNLSSNDSTVDNHNKYNSEYI-----NIQRNKIYELSNKRIRNYK 870
3y 317 PYS-----KLSLEKRIARVPI----- 334
3b 871 LYSNDEIKFYSLEKKYIDNISNNMERVTKYKEMINERKISKMDILYCDKNSLMSGP 930
3y 335 -----SGTGYSTYNAKPE-----VSSIGSLSSNPS 363
3b 931 VLIENNISRENEKNSVYLNNKKNENNFCVGLCHGKGNNDNDNYDGNIRKNEE 990
3y 364 LITSKEISSADGYIFNKDLYEETATAYIRHGDHFYIPKSNIGQPTLPNNSLATPS 423
3b 991 ITKHDEYISREKKNYNSK-----CIRNPDQYK--EDV----- 1022
3y 424 PSLPIPGTSEKHEEDYGPDAIRIADQ-----SGEVMSGHGHNHYFKRD 472
3b 1023 -----LSYHTLDEDKKKNDMMNLIDMNEALIEYNGVYLNILDRKONS--RKD 1071
3y 473 LTBQICAAQKHLB--VYTSNGLDLSLSHEDYGNAKERK-DLAKIEKTAGIMKQ 529
3b 1072 MEKEMEKEMERKEMERKEMERKEMERKEMERKEMERKEMERKEMERKEMERKEMER 1126
3y 530 -----YGVKRESIVYVKEKNALITYPHGDHHPDIDEK-----PVGIGHSNY 574
3b 1127 EINIYKNEIYVNDKLEIYNEEK-LIYPTNES-----DVHNNMNSININCKDQY 1180
3y 575 -----ELKPEGVAKKE-----GNKYVT--GEELIVYVNLK 605
3b 1181 NMLKEVYDNCIAQKKEINIRPLENLKDKYKREYIKNNITIIHNEKRIYQIIN 1240
3y 606 NSTNNQNT-----LANGQKRVSPFPELEKTKIGIMLVKLITPDGVLEKVSQVYG 660
3b 1241 KNPVPIYFNRYENFLIN--HLTYNFK-----NDLEKL-----SYKYSM 1278
3y 661 EGVGN--TANFELDQPY-----LPQO--TFKYTIASDYE--VSDGFTYPTSLAY 707
3b 1279 NNINLVYANHHNNNDYNNKLINQNTLYTLKYVYANLNDNHICKGGGLDYNMNTSK 1338
3y 708 KMASQITFYPPHAGDTYLKVPQAVPKGTALVRFDEFGNAYL-----ENNY 757
3b 1339 ECKNR-----KDKTYL--KIFHYKKKADARFINDSNDYMDIKKYSNDENNY 1389
3y 758 KVGELKLPPLNCGTETKATNKIPVTFMANAYLDNOSTYVEVILEKNOTDKPSILP 817
3b 1390 KLNQ-----KNMISNDBDM--PTLNSEGNFPPSCOP 1422
3y 818 QFRNKAQNSKIDEEVEEPTSEKYEKELSETGNSNSTLEEVPTVPOKVAKFA 877
3b 1423 NLLEK--SYTIDLNYDSMDQTEKYNFNNNN-----DLFNTKRFKN 1468
3y 878 ESYGNKLENYL-----NMDGIELYLPSEGVYKKNMADFGT-----EAPQNGEN 923
3b 1469 FSKGKALFNKKFFVSNEDGYFSF-----KNMILFELKNSNLSLSEYKNSNN 1520
3y 924 KPSENGVSTGVENQPTEN 943

```

DB 1521 CSNNKGDNDIGNMKNANTN 1540

RESULT 3
A:Accession: A54498
major mezozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (15

C:Species: Plasmodium falciparum
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000

C:Accession: A54498
R:Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V.;
Mol. Biochem. Parasitol. 27, 293-302, 1988

A:Title: Variation in the precursor to the major mezozoite surface antigens of Plasmodium
falciparum
Accession: A54498
MUID:88142999

A:Status: Preliminary
A:Molecule type: DNA

A:Residues: 1-1701 <P>
A:Cross-references: GB:M9143; NID:9160412; PID:AAA29653.1; PID:9160413

C:Superfamily: major mezozoite surface antigen
C:Keywords: surface antigen

Query Match 4.2%; Score 226.5; DB 2; Length 1701;
Best Local Similarity 20.2%; Pred. No. 0.0034;
Matches 250; Conservative 175; Mismatches 406; Indels 405; Gaps 66;

DB 1 MRFSTKIYAGSAVIVSLCAVALMOHRSOEN---KNNRYIVDGSOSOSKSENLP 56
DB 500 MRFNNF---DKVVDKIFASRTYVNEKORNNKSSNNSTYV---OKLKALSTYL 553
DB 57 DOVSOKEGIAEQIVIKIDQGVVTSBGHGHYNGKVPYDA---LFSEEL----- 104
DB 554 D-YSLKGI-----SEKDFNHTYTKTGLEADIKKLTIEIKSEKNIIE 596
DB 105 -----LAKDPNTOLKADADIVN-EVKGYYII-KVYG---KYYVYLKDAADNVT----- 149
DB 597 KMFKGITLHSAHNSLEVDIVKLOVQKVLIIKKIEDLKTELEKLNQKDSIHVPITVP 656
DB 150 -----KDEINROKQ-----EHVKONEKYNVNA---VARSQRYTTNDG--- 185
DB 657 QNKPPEYVYIIVLAKKEDKLEFPKVKMDLKEQAVLSSTIOTPLVASE---TTEGSGHS 713
DB 186 -----YVNPADIIEDGNATIVPHGCHYHI---PKDLSASELAAKAHLA 230
DB 714 THTLSGSETETEETEVEET-----VGTTVTITLTPKESAPREKAVVENSIE 765
DB 231 GKNMQSOLSYSTASDNMTQSAVAKSTSPARK-----SENQSLKEL 275
DB 766 HK-----SNDNSQALTKVYLKAKLDEFLTKSYICHKYLIVSSMDQKLEY 812
DB 276 YD-SPGAQRSESDGLVFPDAKISRTPNGVALPHG-----DHYHTIPYSK----- 320
DB 813 YNLTPEEMELKS---CPDLLEFNIGNNIPMAYSLYDSMNIDLOHLEFELIYOKEMITY 868
DB 321 LSLALEK-----IAAWPISTGSGTVST-NAKPNVSSLSGSLSNPSSLTISKLSA 373
DB 869 LHLKLEENHIKLEBOKOITGTSSTSSPGNTVTVAQSTHNSONQOSMSS---TWT 925
DB 374 SDGIYNPDIYVEETAVIVRHGDFHYTPKSNQI-GOPTLPN---NSLATPSPSLP 430
DB 926 QNG-----VAVSSGPAYVEESHDLPLVLSISNDLKGIVSLNLGNKRVPP-LTIS 977
DB 431 GTSHEKHEDGYGFDANRIIAEDSGFVNSHGHNHYFFKCD-----LTEQIKAAQH 484
DB 978 -TEMEKFEYER---ILKNNDIYF---NDIKOFVNSNGVITGLTEQKNALNDE 1024
DB 485 LEEVYKTS-----HN-----GLDLSHEDVYGNAKEMD---LDKITEKTAGIMKV 530
DB 1025 IKKLKDTQLSPDLKYRYKLKIDRLFNKKKEIGQDQKQIKKILLLEQLESKLNLSNPH 1084
DB 531 GYKRE-SIVNKEKNALY-----YPHGDHHDAPIDIEHKPVGI 567
DB 1085 NVALNFSVFPNKKKEELIETENTLENTKILKHGYGVKYTNGE---SPLATISEVSI 1141

DB 568 --GHSHSNTELEKPEEGVAKKEGNKYTTEE-----LTVNVLNKNSTPNNOFTLA 617
DB 1142 QTEEDYANIEKRALSKIDGKNDMLHIGKKTLSLSSGILHLITELK-EVYKNNYNT-- 1198
DB 618 NQOKRVSPFPELEKIGLNLVTL---TTPDGVYLEKVSQKRVGEGVGNIAFELDOP 674
DB 1199 -----GNSPENNKK---VNEALKSYENFLEAVYTVVTF-----PQP 1234
DB 675 YLPQGTFTKI-----ASKDYEVSIDGFTFYPTSLAYVMAQGITFYPHAGDTLARNP 729
DB 1235 DVTPEPLSYRVSAGSSGSTEKEOITPSTG-----SLTLEQOYVOLONTDEEDSLVLP 1288
DB 730 QPAPVPGTVD-----ALVVEDEFHGNALYENNYVGEIKLPIPLNOGTTTACN 779
DB 1289 IFGESDNDDEYLDQVVTGALISVTMDNILSG---FENEYVITLK-PL---AGVYSLAK 1341
DB 780 KIP---VTEMANA-----YLD-----NOSTIYIEVPILEKENOTD 811
DB 1342 QIEKNITFENMLNDILNRLKRRKFFLDVLESIDLMOFHISNENYIIE-DSFKLINSQ 1400
DB 812 KPSILPQFRNK-----AOENSKIDEXY-----EKPISKEKKEKLSGTGS 854
DB 1401 KNTLKSRYKIESEVENDIKFAOEGISTYEKVLAKYKDLSEIKYIKKEKEFPSSPT 1460
DB 855 TSNSTLEVPYVPOEKVAKFA-----ESYGMLENYLFNM----- 891
DB 1461 TTPS-----PAKTEQKSKSKFLPLTNIETLKNLVNKIIDYTLNKKAKINDCENXE 1515
DB 892 -----DGTIELYLPSEGVIRKNNADFTGEAPQNGENKPSENGK-VSTGV 936
DB 1516 AHVKITKLSDLAKIDDKIDLF-----KNTNDFEAIKILINDDKKMLGKLLSTGLV 1567
DB 937 ENQPT-----ENKPADSLPEAPNE-KPYKRENS 963
DB 1568 QNFPNTIISKLEGRQDMINISQHCVRKCCENS 1603

RESULT 4
T30944
surface protein precursor - Enterococcus faecalis
C:Species: Enterococcus faecalis
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
R:Shankar, V.; Baghdadyan, A.S.; Huyke, M.M.; Lindahl, G.; Gilmore, M.S.
Infect. Immun. 67, 193-200, 1999
A:Title: Infection-derived Enterococcus faecalis strains are enriched in esp, a gene
A:Reference number: 220943; MUID:99081742
A:Accession: T30944
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1873 <SHR>
A:Cross-references: EMBL:AF034779; NID:93873186; PID:93873187; PID:AA09858.1

Query Match 4.0%; Score 218; DB 2; Length 1873;
Best Local Similarity 20.8%; Pred. No. 0.011;
Matches 215; Conservative 122; Mismatches 377; Indels 322; Gaps 49;

DB 91 NGKVPDALFSEELMKDNYOLKADIVNEVYGVYIIKVDGRYVYLLDAADAD-----N 146
DB 593 KNLPLDAYSMTKEPDTGNTVDSTGITYVAKNGTFFVDVDFEAVKASQAMENDATYVP 652
DB 147 VTKDEINROKQEHVADNKEVNSNAVAARSQGRYTTNDG---VNPADIIEDTGNAYI 202
DB 653 ITTPEETIOSGKPTGDKDPV-----LANDAFSVLDVYN-----KDFGNASV 695
DB 203 VPHGHIYHIIKSDLSASLAAKAHLAKNNQPSOLSYSTASDNMTQSAVAKSTSKPA 262
DB 696 DANTGIVTTPAKGVSESPITGTT-----PIIVY-----ODGSYGTIDLAIV 738
DB 263 NKENQSLKELDYDPSAO-----RYSESDGLVFPDAKISRTPNGVALPHGDHYH 314

```

Db 739 TVSKN-----IYENPGNIPAGYHKTFTAGEGTSIGSTTVAVADGSLPE-----786
2Y 315 FIPYKSLALEEKIARMPISGTSVSTNAF-PNEVVSLSGLSSNSPSSLTSELSA 373
Db 787 -----DKLPYKAKOG-----YTDKMPPE-----ATOPKADTEFVSSA 822
2Y 374 S-DGIFRKP-----IYETINATYVRKHGDFHITPKNOIGQPTL 414
Db 823 TILDIIEPPGNIPAGYHKTFTAGEGTSIGSTTVAVADG-----VSL 868
2Y 415 PANSIATPSPLIPNGTSHEKEDGCGFDANRIIADESGFVMS-----HGH 464
Db 869 PEDKL-----PYAKADGTTDAKMP-----EATOPKADTEFVSSATKLDIENPGN 919
2Y 465 -----NHFEKPD-----LITEO-----IKAAOHL-----486
Db 920 IPAGYHKTFTAGEGTSIGSTTVAVADGSLPEDKLPYKAKGYTDKMPGEATQPI 979
2Y 487 -----EYKTSHGSLDLSHEDDYPGNAKEMDLPKIEKTAGIMKOYKVRSESTVVK 541
Db 980 KADTEFVSSATKLDKDDADKDPGQKVTLEKKE-PDASEGKINKKDLKDKAKYTK 1038
2Y 542 E-----KNAIIFPGDHHDADPIDEKFPVIGSHSHSYELKPE-BGVAKREGK 590
Db 1039 EKVDSIAGNKKGIVVYYS-D-GSSDEYB--VDYVTDNRSDADKTEFVEGEKVEIGK 1095
2Y 591 VTGEBELTNVNLKNSFTNN-----ONETLANGORVSESPPEEKLGIMLVYL 643
Db 1096 VDLTNVNLPLPGTIVTDTVPGTIDTNPNGVEGIEVETYPD-----GTDTVKV 1149
2Y 644 ---IIPDKVLEKVSAGKTFEGV-----GNANFELDOPVLPQGTFTKTSKYP 691
Db 1150 PVEVTDNRSDADKTEFVEGEKVEIGKVDLTNDNTNL---PLPGGT---TVT-----1197
2Y 692 EYSDGTFTVPLSLAYKMAQOTIFPFHAGDTYLRVNPQFAVPEKGTDLVRFDFHNA 751
Db 1198 DTPGCTIDTNPNGVEGIEV---TY-----PDGKTQVAVPVVVTNDR 1239
2Y 752 YLENNY-----KV---GEIKLP-----IPLANGGT---RTAGNKIPVTMANAYLDN 793
Db 1240 SDADKTEFVEGEKVEIGKVDLTNDNTNLPLPGTIVTDTVPGTIDTNPNGVEGVI 1299
2Y 794 OSTY-----IYVPILEKNOJDKPSILPQFKRKADENSKLDKVEPEKTSKVEK 847
Db 1300 EYVTPDGTQKDTVAVVETVTDNRSDADKTPVVEGEKVEIGKVD-----1343
2Y 848 LSEGNSTNSSTLEVPVDOEKAFAESYGKLENTLVFNODGTIELPLSG--EVI 905
Db 1344 --LNDNTNPLTLPGGTTVTVTPG-----GTIDNTNGTEGVEVTVPGTQKDTV 1393
2Y 906 K-----KNMADFTGEAPQNGENKRPSENGRVS--TGVENQPTENKRPADSLPEAPNEKPV 958
Db 1394 KVPVEVTDNRSDADKTEFVEG--KVEIGKVDLTNDNTNLPT-----LPQGTIVTV 1445
2Y 959 KRENSTONGMKNPEGN 974
Db 1446 TPGGTTIDP---NTPGN 1458

RESULT 5
126868
Major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
Species: Plasmodium falciparum
Date: 19-Nov-1988 #sequence, revision 19-Nov-1988 #text, change 09-Jun-2000
Accession: A26868
Tanabe, K.; Mackay, M.; Goman, M.; Scaife, J.G.
Mol. Biol. 195. 373-387, 1987
Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium
Reference number: A26868; MID: 88011243
Molecule type: DNA
Residues: 1-1701 <TAB>
Superfamily: major merozoite surface antigen

```

```

C:keywords: surface antigen
F:1-19/Domains: signal sequence #status predicted <SIG>
F:20-1701/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 4.0%; Score 215.5; DB 2; Length 1701;
Best Local Similarity 20.1%; Pred. No. 0.012;
Matches 249; Conservative 175; Mismatches 407; Indels 405; Gaps 66;

1 KFSKXYIAGSAVIVSLCAVALNHOHROEN-----KDNNVSYVDSOSSOSSENLTP 56
Db 500 KKFNNF---DEVDYKIFSAITTYNEKORINKNFSSNSNVYV---QKIKLSTYLE 553
2Y 57 DOVSOKEGIOEQIYIKITDQGYVTSBGDHYHYNGKVPYDA---LFEEL-----104
Db 554 D-YSLRKG-----SEDFNHYIYLLKGLLEADIKLKEIKSSNKILE 596
2Y 105 ---LMDPNYOLKADIVN-EYKGYIT-KVDG--KYIYLDAAHADVRT-----149
Db 597 KNFKGLTSHANASLEVSIVKIQVOKVLLIKIEDLRKIELEKNAOLKDSIHVPNIYKP 656
2Y 150 -----KDEINROK-----EHVNDNEKVNNA---VAASGRTYNDG---185
Db 657 ONKPEPYIYLLKKEVDKLEPIPKVMDLKKKQAVLSITQPLVASE---TDEGSH 713
2Y 186 -----YFNPADIIEDTGNAYIVPHGHYHI-----PKSLSASEIAAKAHIA 230
Db 714 THLSQGETEVEETEETET---VGHYTVITLIPKRESAPKEVKYENSIE 765
2Y 231 GRNNQPSQLSTSSSTRSDNNNTOSVAKGTSKPAK-----SENLSLKL 275
Db 766 HK-----SNDNSQALRTYVLLKLDELTRSYCHKIYIVSNSMOOKLEY 812
2Y 276 YD-SPSAORYESDGLVDPKAIISRTPNGAIPG-----DHYHFIYSK-----320
Db 813 YNLPPEEKELKS---CDPLDLFNIONNIPAMTSLYDSANNDLQHLFEELYOKEMITY 868
2Y 321 LSALEK-----IARMPISGTSVST-NAKPNVVSLSGLSSNSPSSLTSELSA 373
Db 869 LHLKLEENHIKLLKEQKQITGSTSFGNTVYTAQSAHNSNQNSNASS---TNT 925
2Y 374 SDGYTFNPKDYEETATAYTRHGDHFHYIKSNQI-GQPLPN--NSLATPSSLPINP 430
Db 926 QNG-----VAVSGPVAVESHPDLTVLSISDLKGIIVSLMNGTKVPPN-LTIST 977
2Y 431 GTSHEKHEDEGFDANRIIADESGFVMSGDNHNEFFKD-----LTEQIKAAQKH 484
Db 978 -TEMEKTYEN-----ILKNNDYF---NDIKQPVKNSKVITGLTETQKNALNDE 1024
2Y 485 LEEVYTS-----HN-----GLDLSHEDDYPGNANEMKD---LDKIEEKIAGIMKQY 530
Db 1025 IKKIKDPLQLSFDLYNNYKLLKDLPLFKKKKELGDDKQKQIKLTLTKQLLEKSLNPNH 1084
2Y 531 GYARE-SIVYKKEKNAIL-----YRGGDHHDADIDEKFPVGI 567
Db 1085 NYLQNFVTFVKKKKEAEIETENTLENTKILKHKGIVKYKNGE---SSPLKLTSEESI 1141
2Y 568 ---GSHSNYELFKPEEGYAKKEGKVTGEE-----LNVVNLKNSFNNOPTLA 617
Db 1142 QTEBNYANLEKFRALSITDGLNDNLHKKKLSFLSSGHLHLEIK-EYIKANKNT--1198
2Y 618 NGOKRVSPSPPELEKLGIMLVYL---ITPDGKVLKVSAGKVGSGVGNIANFELDOP 674
Db 1199 -----GNSPSENNK--VNEALKSYENFPPEAKVTVVTP-----POP 1234
2Y 675 YLPGQTKYIT-----ASKDYPEYSYDGTTPVSLAYKMAQOTIFPFHAGDTYLRVNP 729
Db 1235 DVPSPSLVSVSVSGSGSTKETQIPTSG---SLTLEQVQVQLONYDEEDSLVYLP 1288
2Y 730 QFVAVPGTD-----ALVYFDEFHGNAYLENNYKYGKELPLPKNOSTTTAGN 779
Db 1289 IFGESEDNEDYLDQVVTGALISVTMDNITSG--PENETDVIYIK-PL---AGVYSLAK 1341

```

```

780 KIP---VFMANA-----YLD-----NOSTYVEVPLEKENOTD 811
1342 QIEKNTITFNLMNDILNRLKRRKRYFLDVLDESLOMOKHSSNEYIE-DSFKILNSEQ 1400
812 KPSILPOKRNK-----AOENSKLDERV-----DEPTSEKEREKLSGTNS 854
1401 KNTLKSYKIKESVENDIKFAQEGISYEYKVLAKYKDLSEIKKIVKEKEKPPSSPT 1460
855 TSNSTLEVPYVDVQEVAFPA-----ESYGLKLNVLNFM----- 891
1461 TTPS-----PAKTDEQKESKFPFLNIEFLYNNLVAKIDYDYLNLAKINDCNVERDE 1515
892 -----DGTIELBSGEVIRKNNADPTGEAPOGNGENKPSENGK-VSTGTV 936
1516 AHAYITKLSLDAKIDDKIDLE-----KFNDFEAKIKLINDTKDKMDGKLSLSTLY 1567
937 ENQPT-----ENKPADSLPEAPNE--KPYKPPNS 963
1568 QIFPNTIISKLEGGFQDMNLNISOHCYKCOCPENS 1603

RESULT 6
18372
Repeat organelle protein - Plasmodium chabaudi
Species: Plasmodium chabaudi
Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
Accession: T18372
Werner, E.B.; Taylor, W.R.; Holder, A.A.
Mol. Biochem. Parasitol. 94: 185-196, 1998
Title: A Plasmodium chabaudi protein contains a repetitive region with a predicted spe
Reference number: Z18922; MIMD:98418765
Accession: T18372
Status: preliminary; translated from GE/EMBL/DBD
Molecule type: DNA
Residues: 1-1939 <WEB>
Cross-references: EMBL:U04145; MID:91151157; PID:91151158; PIDN:AAC63403.1

Query Match. 4.08; Score 214; DB 2; Length 1939;
Best Local Similarity 17.98; Pred. No. 0.017;
Matches 220; Conservative 183; Mismatches 415; Indels 408; Gaps 53;

28 HRSQENKNNRVSYDGSQSSQKSENLPDVSQKGIQARQIVIKIDQGVYVSHGHY 87
462 HKNELNELNOLSKLNKKNKNEN-----TELNDKI 494
88 HYNGKPYDALFEELMKDPNTOLKDA---DIVNEKGGYIIKVDGKYYVLKDAH 143
495 SLSNSEV-----NILNKDQGLNDIKTLNDLNLK----- 526
144 ADVNRTK-EINROKQEHVKNENKVNVAARSGRTTNDGYF-----NPADII 194
527 -NEINTSDNKKMKKEDLAMEEKGCVVIDETEKYKKEITPMLKEKLEKEKYADLN 585
195 EDTG---NAYIVPBGHHTYIPKSDLSASELAAKHAHLAKNMPOLSSTASD--- 247
586 DEISLRKSTYKE-----KEPIEKKEFEYKINLFKNKPEKKIITENELSLRK 637
248 -NNTQSVAKGSTSRPANKSENIQSLIKELY-----DSPAQRYSESDBGYF 292
638 YDNBQGLIKQDELNIQKLEKEKYL-QLYDNMNMFRSCTKIDMPSYSENIKSDLDVDF 696
293 DPAAIISTPNGV-AIRPGDHYHTPYSK-----LSALEKIANMVPISGTG----- 338
697 VTAYIKRDESSDANPTTKEKVAELERNAIVALEEKHEEIKLQEGHKEVYLR 756
339 -----STVSTNKPPEVSSLG-----SLSSNPSSLTTS----- 367
757 LGBOHKEETIILEKHKOVTYKLGBOHKEKNTIKLEEKHKOVTYKLGDOYKEELAKKEH 816
368 -----KELSSADGYIFNPKDIVEETATAYIVRHGDHFYIFKSNIGQPTLPN----- 416
817 AYVVAELEEKHKLQEGH-----KENVDELEK-----RHADFVGLGEKHKAETAKLEGHKS 868

```

```

OY 417 -----NSLAPSPSLPINPCTSH-----EKEEDGYGFANDRIIADES 455
DB 869 EMNEVERHADPVBGLEKKAETAKLQEGHREYVAGLEENKKE-----YVAELE 919
OY 456 GFVASHGDHNYFFPKDLTEBOIRA--AOKHLEEKYSHNGISLSHEDYPGNANEMK 513
DB 920 -----KHKEIKALKEGHKEVMAELGEKKEVYA-----GLEKHNLEBGHKEVVALE 968
OY 514 -----DLKITEK-----IAGIMQYKRSIYVNEKKAIIYPHGH 554
DB 969 KRHADVAVLEEQHKAELIKLGEHKEVYVAGIEKY--KVAIKLAEHKOVTYKLGBOH 1036
OY 555 H---ADPIDEKPY-----GIGSHSN-----YELFKPEE 581
DB 1027 KEETAKLEDGHEVYNEVEKKNASILNLENNKEMKIKLEEKRESADLVKTIQODE 1086
OY 582 GVARKKGNVYTGELTYVNMILKNSTFNNQNTLANQOKVSEFPPELEK-KLGINML 640
DB 1087 EV-KNSNKKI---ELFTVINDLSIMCYKKOILEVEKERNEYR--EINKIKIYQNM 1140
OY 641 VKLITPDQVLEKYSG-KVGEQGNINAFELDOPYLPQGFKTITASKDYPEVSYDGF 699
DB 1141 KDM--NDKKILEKENEIKKIKKLSNYKVFETKE-----NTYK----- 1176
OY 700 TVPTSLAVKMASQTIFYPFHAGDYTLRVNPOFAVPKGTDALY--RVDFEFGNAYLENNY 757
DB 1177 -----NSEVYVNEKERRIIVDSVCENISESDVEG-- 1206
OY 758 KVGEEK--LPRIKLNQGTTRPAGNKIPYTFANAYVLDNOSTYVEVPLEKENOTDKPSI 815
DB 1207 KGNMLKMTLSLKKERNFTSINDKNSESELVDLT--KSAVINKEIEMKKEIENGKNI 1263
OY 816 -----LPQFRKKAQENSKL-----DEVBEKTSKYE----- 844
DB 1264 EDLNKIKILDSNELINENMKNVLDENNKLKEIEIDNKLNEKENENELIANDDI 1323
OY 845 ---KEKSEGTNSTNSNLEEMPTVDVQEVAFKAEYKGLNVL--FN-----M 891
DB 1334 IKLKEISEMKDEBEKLTKEKI-----KLKNDIEQNKETKIKENLAKFENENETSL 1379
OY 892 DGTIELYLPSEGVIRKNNADPTGEAPOGNGENKPSENGKYSTGVENOPTENKPADSLPE 951
DB 1380 KNOIEIEKMKLELNKNVKNELLAEKERTNMSISDNKIVENNILED--TDSK----- 1430
OY 952 APNEKPYABENSTDGMLNPGNNGSDPMLDPALEAPVDPVQEKLEKFTASVGLG-- 1009
DB 1431 -----QNNLNK--NVEDKGTGDDINCKRNNDQAKEISYLDDEIKIKISMLGEEELNR 1478
OY 1010 -----DSVLENNMG---TELRLPGE 1028
DB 1479 KNSTDEKYNKLTNELKELKLNKKE 1504

RESULT 7
A42404
Collagen adhesin - Staphylococcus aureus
C:Species: Staphylococcus aureus
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 10-Nov-1995
C:Accession: A42404; S27665
R:Patl. J.M.; Jonsson, H.; Guss, B.; Svitalaki, L.M.; Wlberg, K.; Lindberg, M.; Hook
J. Biol. Chem. 267: 4766-4772, 1992
A:Title: Molecular characterization and expression of a gene encoding a Staphylococcus
A:Reference number: A42404; MIMD:92165839
A:Contents: FDA 574
A:Accession: A42404
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1185 <PAT>
A:Cross-references: EMBL:M81736
A>Note: sequence extracted from NCBI backbone (NCBIF:83982)

```

Query Match 3.8%; Score 205.5; DB 2; Length 1185;
 Best Local Similarity 19.7%; Pred. No. 0.021;
 Matches 214; Conservative 139; Mismatches 411; Indels 323; Gaps 50;

```

36 NNRVSYDGSQSSQSKSENLT-PPQVSGKEGIQAEQIYKITTQGYTSHGDHGHYNGK- 93
b 197 NNEKSYV-----SDITIKDIOGGQOLDISTLININT-----GTHSYSGOS 240
y 94 --VPYDLFSEELMKQPNQOLKMDADI-----VNEVKGGIYIKVDS--KIYYYLKDA 142
b 241 AITDFEKAFFGSKITVYDNTKNTKIDVTIPQIGSYNSFSINTKIKITNEQKKEFNNSQAM 300
y 143 HADNVRTKDEINROKOEHVNDNEKVSNNV-----AVASQGRATTN 183
b 301 YQEH--GKEEVNCKSFPHYHTNINANGIGTYKGLKVLKQKDKTAPLANVAFKLSK 358
y 184 DGTYF---NPADIEDTGNAYIPHGGHYHYPKSDLSASELAANAHLAKKMQPSOL 239
b 359 DGSVYKDNQKEIEITDANGI-----ANIKALPSGDIILKEIEAPPTFDKDE---- 408
y 240 SYSTASDNTQSAKSTSKPANKSENLOSILKELTDSPSAQRYSQGLYTPDPAKIIS 299
b 409 -YPTMKOTNOGTF--TTENAKAIKTKDY-----SAQR-----VWGTOVK 450
y 300 RTPNGVAIPRGDHYHFIYKSK--LSALEEKIARVAPISGTYSTVNAKPNVYVSSLSGL 357
b 451 PTIFYKLYKQDDNQNTTPVOKAEIKKLED-----GTTKVTWMSLTPEND----- 493
y 358 SSNPSLTTSKELSSADGTYIPNKOIYEETAAIYIRHGDHYHYSK--NOIGQPTLP- 415
b 494 -----NNSLATPSPSLPI-NPQTSHEKEHEDGYGFQDANRIIADSEGFVSHGDNHYTF 516
y 416 -----NNSLATPSPSLPI-NPQTSHEKEHEDGYGFQDANRIIADSEGFVSHGDNHYTF 469
b 517 GYTKKEGLVYNTNEKRIETTSISGEKVMDDKQNDQKR--PEVSYNLANG----- 567
y 470 KQDTEQIKAQKHLEEVYTSANGDLSLSEHDYDGNAKEMKDKITE-EKIAGIMK 528
b 568 -----EKVYK--LDVTSETMKY-----EFDLPRKYEGKRIEYTVT 602
y 529 QYQVK-----RESIYVVK--EKNALIYPHGDHHAADIDEKPKPGIGHSHSNVLEFK 578
b 603 EDHVKDVTDTINGNTTINKYTPGETSATVTKNMDDNNQ--DGKRPTEI-----KVELY- 654
y 579 PEEGVA-----KREGKRV-YTGELEFNVMVLKNSPNNQNT 615
b 655 -QQKATGKATLINESNNHTWTGDEKAKAGQAVKTYVEELFY--KGYTHVDND 709
y 616 LANGQKRVSEFPPELEKLGINKLVKITPDGKVLKESVSGYFEGE-----VGNIAN 668
b 710 M--GNLIVTKKYPETTSISGEKVMDDKQNDQKRPEKVSNNLANGDEKVTLDVSETN 767
y 669 FELDQPLP---GQTKYTIAS--KDYPEVSDIGTFYPTSLAYMAAQITFYPPHAG 721
b 768 WKTEFKDLPKYDESKKLEYVTEDEHVDY--TTDINGT-----TTNKTPG 812
y 722 DTYLRV-----NPOFAPRGTDALVFPDE-----FHGNAYLENNYK 759
b 813 ETGATYTKKNDNNQDKRPTKELVYDGAATGTALINESNNHTWTGIDEKAKG 872
y 760 GEIKLPLPKLNQ--GTRTAGKIPVTFMANAYLDNOSTYIYEVPLLEKENGTDK--- 812
b 873 QQVAYVEEELTKVGYTHVDNDMGMLIYTNKYTPETTSISGEKVMDDKQNDQKRPEK 932
y 813 --PSILPQFRKNA-----QENSKIDE--KVEPEKTSVEVEKEKSEIGNST 855
b 933 VSVNLANGKRVKTLDTVSETNKTPEKDKLPKDEGKRIYTYTEHVKYXTDINDGTTI 992
y 856 SNSTL--EEVPTVD-----PVQEKYAKFESGKMLKENTLFPNDGTYELYL 900
b 993 TTKYTPETTSATYTKKNDNNQDKRPTKELVYDGAATGTALINESNNHTWTG 1052
y 901 SGEYIKKMAADFT-GEAPQNGENKRPSENGKSTGVLENQPTENKRPADSL-PEAPANEK-- 956

```

Db 1053 LDKRAGQOQKYTVDELTKVNGYTHVDNDMGMLIYTNKYTPKPKPKPKPKPK 1112
 QY 957 PVKPENS 963
 Db 1113 PTPKPHS 1119

RESULT 8

SAQOM

Major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
 N:Alternate names: 195K glycoprotein
 C:Species: Plasmodium falciparum
 C:Date: 30-Sep-1987 #sequence_revision 31-Mar-1991 #text_change 09-Jun-2000
 C:Accession: A23386; S06361

R:Weber, J.L.; Leidinger, W.M.; Lyon, J.A.
 Nucleic Acids Res. 14, 3311-3323, 1986

A:Title: Variation in the gene encoding a major merozoite surface antigen of the huma
 A:Reference number: A23386; M0ID:86205236
 A:Accession: A23386

A:Molecule type: DNA
 A:Residues: 1-1104 <WEB1>

A:Cross-references: EMBL:X03831
 R:Weber, J.L.; Sim, B.K.L.; Lyon, J.A.; Wolff, R.
 Nucleic Acids Res. 16, 1206, 1988

A:Title: Merozoite surface protein sequence from the Camp strain of the human malaria
 A:Reference number: S06361; M0ID:88143999
 A:Accession: S06361

A:Molecule type: DNA
 A:Residues: 1104-1726 <WEB2>

A:Cross-references: EMBL:X03831
 A:Comment: The merozoite stages of different strains have strain-specific surface ant
 C:Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The
 C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat

F:1-19/Domain: signal sequence #status predicted <Sig>
 F:20-1126/Product: major merozoite surface antigen #status predicted <Mat>
 F:67-87,91-96,100-105,109-120/Region: 3-residue repeats (S-G-T)
 F:757-765/Region: 3-residue repeats (T-E-E)
 F:133,272,501,567,638,827,839,924,944,990,1016,1114,1221,1613,1658/Binding site: carb

Query Match 3.8%; Score 205.5; DB 1; Length 1726;
 Best Local Similarity 20.0%; Pred. No. 0.038;
 Matches 246; Conservative 172; Mismatches 414; Indels 397; Gaps 65;

```

1 MKTSKTIAGSNVYISLSICAVALNHRSGEN-----KONNRVSYDGSQSSQSKSENLT 56
b 531 MKFNNEF--DKDVDRIFSAITYNEKORYNNKSSNNSYNV--OKLKAUSTYLE 584
y 57 DOVSQKEGIAQEIIVIKITQGYTSHGDHYHYNGKVPYDA--LESEEL----- 104
b 585 D-YSLRGT-----SEKDNHYITLTKGLEADIKKTEIKSENNILE 627
y 105 -----LAKDNYOLKDAIVN-EVKGTYI-KYDG--KIYYVLKDAHADNVRT----- 149
b 628 KNRKGLTHSANAILEVYDIQVOKVLLIKKIEDLKIELFLKMQKDSIHVPNIYK 687
y 150 -----KDEINROQ-----EHYKDNKVSNNV--VARSGRYTNDGYF 188
b 688 QNRPEPYLLIVLKEVDKLEFIPKVMKKEQAVLSITQPLVASE--TTEDG--- 741
y 189 NPADIEDTGNAYIPHGGHYHYHYSK-----DLSASELAAKAHAKKMQPSOLSY 241
b 742 -----GHTHTLSQSGTEVEETPEETETGYGHTTIVTILPKREVY 784
y 242 SSTA---SDNNTQSAKSTSKPANK-----SENLOSILKEIYD-SPSA 281
b 785 VENSIEHKSNDNSQALTKYVLLKLEDFLTKSYCHKIYLVSSNSMDQKILEYNTLPEE 844
y 282 QRSSESDGLVTPAKIISRPNGVAIPHG-----DHVHPIYKSK--LSALEEK 327
b 845 ENELKS---CDLLELFIQNIQNNIPMYLSLYDSMNDLOHLEFELYQKEKITYLHKLKE 900

```


328 -----IARVPISGCTGYST-NAKPNVSSLSGSSNPSLITTSKELSSASDGIYN 380
 b 901 NHIKKLEEQKQITGISTSSPGNTVNTAQSATHSNQOQSSASS---TNGQNG--- 953
 y 381 PKDVEETATAYIYRAGDHHYIPKSNQI-GQPTLPN-NSLATPSPSLDIPNPTSEHK 437
 b 954 ----VAVSSGAVVSESDPLVLTSLSDKGYSLMLGKTKVYPNP-LTIST-TEHEKF 1008
 y 438 EEDQYGDANRIIADESGFVMSHGDHNYFFKND-----LTEEQIKAQKLEEVKTS 491
 b 1009 YEN-----ILKNNDTYF---NNDIKQFVKSNSKYITGLTFETQKALNDEIKKLDI 1056
 y 492 -----HN-----GLDLSHEDDYPGNAKEMD---LDKIEKINGIMKQYVKE-S 536
 b 1057 LQSLFDLYNKYKILDLIFNKKKELGODKQIKKTLTKDLKESKLSLNNPANHVLNFS 1116
 y 537 IYVKEKNATY-----YPHGDHHDADIDEHKPVGI--GSHS 572
 b 1117 VEPKRRKEAEIETENTLENTKILKHYKGYVKNCE---SPPLKLSSEVSIOTEDNYA 1173
 y 573 NYELFKPEGVAKKGNKYVTGEE-----LTNVNLLKNSTFNQNFLLANGQKRV 624
 b 1174 NLEKRYLSKIDGLNMLHGLKRLSFLSGHLHLETK-EVIRKKNY----- 1223
 y 625 FSPPELEKLGIMLVKL---IIPDGKYLEKSGKYPGEGVGNANPELDOPYLPQTF 681
 b 1224 GNSPSENNK--VNEAKSYENFLPEAKVTVTP-----PDPVTPSPL 1266
 y 682 KYTT-----ASKDPEVSYDQTFVPTSLAVKMAQOTLFYFPHAGDTYLRYNPOFAVPG 736
 b 1267 SVRSGSGSKETQIPTS-----SLTELOQVQVQVQVDEEDSLVLPFPGESD 1320
 y 737 TD-----ALVRYVDFEFGNAVLENNTKVEIKLPKLNQGTTRAGNKP--V 783
 b 1321 NDELDQVVTGEALSTVMDNLISG--FENERDYIYK-PL---AGYRSIKQIEKNIF 1373
 y 784 TFMANA-----YLD-----NOSTIYVPILEKENTDIPSLIPQ 818
 b 1374 TPNLNDILNLSRLKRRKRYFLDYLESDLMQFKHISSEYITE-DSFKILNSEQNTLLKS 1432
 y 819 EKRNR-----AOENSKLDEKY-----EKPTEKEREKLETSNSTNSTLE 861
 b 1433 YKTKESVENDIKRQGISYIEKVLAKYKDDLESIKKVIKEKEKPPSSPTTPPS--- 1489
 y 862 EYPTVDVQEVAKFA-----ESTYKLELVLFM----- 891
 b 1490 --PAKTDQKKESFPLFLNIETLYNNLVNKKIDVILINLAKINDCNVEKDEAHVKITK 1547
 y 892 -----DGTIELYLPBGGEVYKRNADFTGEAPGNGENKPSNGK-VSTGTVNOPT-- 941
 b 1548 LSDKAIIDKIDLF-----KNHNDFEAIKKLINDTKMDKGLSTGLVQNFPTI 1599
 y 942 -----ENKPADSLPEAPNE--KPYKPN 963
 b 1600 ISKLEKFPQDMNISQHCYKQCPENS 1628

RESULT 9
 41859
 GA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae
 Species: Haemophilus influenzae
 Variety: strain HK613
 Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
 Accession: C41859
 Poulsen, K.; Reinholdt, J.; Killian, M.
 Bacteriol. 174, 2913-2921, 1992
 Title: A comparative genetic study of serologically distinct Haemophilus influenzae ty
 Reference number: A41859; M01D:9234949
 Accession: C41859
 Status: preliminary; not compared with conceptual translation
 Molecule type: nucleic acid
 Residues: 1-1849 <PDB>

A: Experimental source: strain HK613
 A: Note: sequence extracted from NCBI backbone (NCBI:97285)
 C: Superfamily: IGA-specific metalloendopeptidase
 C: Keywords: hydrolase; metalloproteinase

Query Match 3.8%; Score 203.5; DB 2; Length 1849;
 Best Local Similarity 19.3%; Pred. No. 0.054;
 Matches 239; Conservative 153; Mismatches 449; Indels 397; Gaps 59;

28 HRSQENKNNRVSYDSSQSSQSENLTPQVQSEIQ-----AEQVIKLTQ 77
 b 125 HRDVSSENR---YTYEKNFPTENTVSTFKEEQDAQKREDYMPRLDKFTVEVAPI 181
 y 78 GYTSHGDDHYHYNGKPYDALF-----SEELMKDNYOL---KQAD---YNEVKG 125
 b 182 EASTANNKGE-YNNSDKYPLFVRLGSGQFYKKSRYOLITLTKQGNLNNMDVYG 240
 y 126 YIRVDGKYVYLLKDAHADNVRTKDEINROKQEHVADNEKVSNNVAVSQGRYTTNDG 185
 b 241 DNLELVGNATY-----GIAGTPYKNHEN-----NGLIGFNSKEHSDPKG 283
 y 186 YV-----FNPADLIEDTGNATYV--PHGGHYIYKSDL-----SASELAALAAHLAK 232
 b 284 ILSDPLNTYAVLGDGSPLEFYDREKGMFLGSDYFMAQYNNKSSQWENNYIYHEPAK 343
 y 233 NMPSQSLYS--TASDNTQSV--AKGSTSPARKSENLSLKELYDSPSAORY----- 284
 b 344 IYQ-----QYSAGSLGSTQYQWATQSTSTYTGCGEPLSDLDGDKDPHGSITLKG 399
 y 285 -----SESDGLVFPDAKILSRPN-----GVALPHGD---HYHFLPYSKLS 322
 b 400 SGTITLNNHIDGAGCLFEEDYEVKGTSDSTWKGAGVSADKTYWKNYHNPYDLA 459
 y 323 -----ALEEK-----IARVPISGCTGYSTYNAK- 346
 b 460 KIGKGLVVEGKGNKGLKVGDTVILKQKADANNKVOASQVIGYSTLTVLNDKQ 519
 y 347 --PREVYSSL--GSLSNPSL-----TSKELSSADGYIYFKD 383
 b 520 VDPNSTIYFGRGRDLNGLNSLTDHIRNIDGARYVNNHNTNTSNITIGESLTPN-- 577
 y 384 IVEETATAYIYRAGDHF-----YIPKSNQIGPTLPNNSLATPSPSLIPNG 431
 b 578 --NTITSYNLEADDDHPLRISIPYKQLYFNDNR-SYTLKKG--ASRSLSPQNSG 631
 y 432 TSHKHEDEGYFD-----ANRIIADESGFVMSHGDHNYFFKNDLTFEQIQAQKHL 485
 b 632 ESNENWLYMGRTSDKARNVNHINERNNGF-----NGYF-----GEETKATON-- 677
 y 486 EYVTSNGLDLSHEDDY--PGNAKEMDLDKIEKINGIMKQYVKESTIVNKEK 543
 b 678 GKLVNTFG--KSDNRFLLTGTNLNGDL-----NVEK 709
 y 544 NATIYPHGDHHDADIDEHKPVGI-----HSHSNYELFKPEGVAK-----EGN 589
 b 710 GTPLSGRPYRHADI-----AGISSTKDPHFTENNRYVEDDMINNFATIMNTGN 764
 y 590 -KYVTGEELTNVNNLKNSTFNQNFLLANGQKRVSPPELEKLGIMLVK----- 642
 b 765 ASLVSGR--NVANITSNITASN-----NAQVHIGY-----KTGDYCVASDYTGX 807
 y 643 LITDGYLVEVSKYVGEV-GNI-----ANPELDOPYLPQGFYKTIYAKDVPESYD 696
 b 808 VTCHNSNLSERKALNSFPTNLGRNVNLTENASFLIGKANLFG-----TISIGISQVA-- 860
 y 697 GTFVPTSLAVKMAQOTLFYFPHAGDTYLRYNPOFAVPGTDALVRYFDEHGNNA----- 751
 b 861 ----LKENSRLHGLGNSVNVQNLNTNGHILHANAQNDANKYTYVTTLVYNSLGSNGSYTW 916
 y 752 -----YLENNRYVGEIKLPI-----PKLNGT---TTPAGKLPVTFMANAYL 791
 b 917 VDFNNNSNKRYVVKATGNTLLQVADKTGEPNHNELTLFDASVATNNLEVT-LANGSV 975

A:Cross-references: EMBL:M59706; NID:g160311; PID:g160312
 A>Note: sequence extracted from NCBI backbone (NCBIN:77801, NCBI:P:77802)

Query Match 3.6%; Score 196.5; DB 2; Length 1271;
 Best Local Similarity 19.1%; Pred. No. 0.067;
 Matches 219; Conservative 163; Mismatches 434; Indels 333; Gaps 55;

19 SLCAVYALNDRSGENKNNRNVSYVDSSQSSQSENLTPQVSKREGIQAMQIYIKITDGS 78
 153 SLSDLELLENSSDNDLKDPTIS--TEPPPKKKKLDQD-----LNDPL-----EP 197
 79 YVSHGDHYHYNGKVPYDALFSEELMKDPNTOLKADIVNEVYKGGYIKVDGKYYVL 138
 198 FPPQIKHDIYKKN-----LINEDESEPPPKKKKVDNHNNEKN-----VFH 239
 139 KDAAHADN--VTKDEINRQKQEHVKNKNSNAVAR-----SGRGYTN-DGYV 187
 240 ENGSANGNQSILKSPDEHLKDEKTEENELVHENLSTPNDPIEQILNDPEQETNIOEL 299
 188 FNPADITEDTGNAV-----YVHGCHYHITPKSDLSASGLAAKAAKHLAKGN 234
 300 YNEKQNEEKNQSPISLDKEPTNEDILPNHNPLENIKQSE--SEINHYODHALPKEN 356
 235 QPSQLSYSSASDNNQSVAKGSTSKPANKSENLQSLKELYDPSAQRYSQSDGLVFPD 294
 357 TIDL-----DKQKHIDQSOHINIVLOENNINNHQLEPOEKPNIES-----FEP 401
 295 AKIISR--TPNGVA-----IP--HGDIYHIFPYKSLALEEKLARMPVIGSTGTVS 342
 402 KNIDSEILLENTVEETIIDVSPKSNHETFEETSESEHEEVA--KNAH 453
 343 TNAPKNEVSSLSLSSNPSSLTTSKLSASQGYITNPDIYEBETATAY-----IV 394
 454 EYVEHEETVSO--ESNPEKADNDGNVSONSNNEI--NENEFVSESEKHEHAENESSL 508
 395 RHGDHFIYIKSN--QIGQFTLIPNNSLATPSPSLIPNGTSHEKHEHD-----GYGFA 446
 509 EEGHHEIYVEQNNNEESGKLVN-----DEGFEENHNESESSEVSNSELINE 557
 447 NRILAEDESGFVSHGDHNYFFPKDLTE--EQILAAKHLEEVTSINGDLSLS-- 500
 558 NEVESEKKS--VTEPAHEHEVSESENPPEAHEESSIEHAQHEIYVPENDDESGSGL 615
 501 -----HEODIPG-----NAKEKDKDKIEEKIAGI--MKQYGVKRESIV 538
 616 VDRNEGDPEEPNEEPEPDNDSELSENEIYSESEKSYSEPAHEIYSEKSESEPAHEV 675
 539 VNEKKNALIPPHDHHADPIDEHKPVGISHSNYELFPREGVAKKEG--NKVYTG 595
 676 IYSEKST-----SEPAHEVES--EOSNNE--PSE--KKDPVPSKPEET 716
 596 ELTNVVMILKSTNNQNF--TLANGQKRVSPF--PPELEKKIGIMLVKLTTPD 647
 717 EKVYDQPKIYDILIEENFVDSQNPQEPPEPVKIEKVPSEKNKASV-----D 767
 648 GAVLEKVSQVFGEGGNANFELDQYLPQGTFFKTTIASKDYPERYSYDGTFFVPPSLAY 707
 768 PEVKEK-----ENVSVEVEEK-----QNSQSEVEEIPVNEDEFEVEVHEEQDLIDH 812
 708 KMASQTFYFFHAGDYLVRNPOFAPVKGTDALVRVDFEHGNAVLENNKVGKILPIP 767
 813 K-----YDPEIIV-----EVEIPEISELHEEVA--HEIYIEIEVEYFP 847
 768 KLNQGTTRTAGNKIPYTFANAYLNDQSTY--IVEV--PILEKENDTKR-----PS 814
 848 EPNNQNNFQGINE-----DKSAHIQHIEIYVEEILPEDDKNEKVEHEIYVEE 896
 815 ILPQFKRNKAQ--ENSKIDVEVEEPPKSEKYE-----KEKLSIGNSTSSTLEEVTVVD 867
 897 ILPEDKNEKQHEIYVEEILPEDDKNEKVEHEIYVEEILPEDKNEKQGOHEIYVEEIL 956
 868 PVQEKVAKFASGYKMKLENTL-----FNNDG-----TILLYLP--SGEVIKKNMADFTGEAP 917

DB 957 P-EDKNEK-VEHEIYVEEILPEDKNEKQHEIYVEEILPEDKNEKQHEIYVEEILP 1014
 QY 918 QGNGENKPSNGKSVSTGYENOPTENK-----PADS----- 948
 DB 1015 EDKNEKQHEIYVEE--EILPEDKNEKQHEIYVEEILPEDKNEKQHEIYVEEIL 1070
 QY 949 LPEAPNPK-----PVKPESTNDQMLPBGVNSGDPMLDPALEE 987
 DB 1071 LPEDKNEKQHEIYVEEILPEIYVEEIVPSPQNTNNNETITKEPKKNEFSVEEKAIPQ 1130
 QY 988 APAPDPVQE 996
 DB 1131 EPVPTLINE 1139

RESULT 12
 F82884
 hypothetical protein U0495 [imported] - Ureaplasma urealyticum
 C:Species: Ureaplasma urealyticum
 C:date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: F82884
 R:Glaas, J.I.; Lefkowitz, E.U.; Glaas, J.S.; Helner, C.R.; Chen, E.Y.; Cassell, G.H.
 submitted to GenBank, February 2000
 A:Description: The complete sequence of Ureaplasma urealyticum. Alternate views of
 A:Reference number: A82870
 A:Accession: F82884
 A:Status: preliminary
 A:molecule type: DNA
 A:Residues: 1-5005 <GLA>
 A:Cross-references: GB:A002148; GB:AF222894; NID:g6899495; PIDN:AAF30907.1; GSPDB:GN
 A:Experimental source: serovar 3; biovar 1
 C:Genetics:
 A:gene: U0495
 A:genetic code: SGC3

Query Match 3.6%; Score 196; DB 2; Length 5005;
 Best Local Similarity 17.3%; Pred. No. 0.59;
 Matches 234; Conservative 177; Mismatches 454; Indels 490; Gaps 57;

6 KYIAASAVIVS-----LSLCAVALNDRSGENKNNRNVSYVDSSQSSQSENLTPD 57
 38 KVEEGEDVYVNVQDVGNLIKIPALVNNKHQLDLSNN-----PNETKD 83
 DB 58 QVSQKEGIAEQIYIKITDQGYTSHGDIHYNGKVPYDALPSE-- 103
 84 KIYHFDISIE-----DEYH-----IPKETIISDQVDSVTLNNEVD 122
 QY 104 -----LAKDPN-YOLKADIV-----NE-VKGGYIIVDGKYYV----- 136
 DB 123 LGRKIIYKNSVIONKPTTILTVNDPKLALNERKLTITLALNDRPKHCVSAGLDIYK 182
 137 -----YLKDAHADNVRTDEINRQKQEHVKNKNSNAVARSGRY 180
 DB 183 NSLVPENHLQTFAYIKLVFKNNNLNLTDLNAEHSQTKIN--LIKQYIGYFRIF 240
 QY 181 TYNDDGYVFNADITIEDTGNAVIVPHGCHYHITPKSDLSASGLAAKAAKHLAKGNQSO 240
 DB 241 STN-----TLVLED-----YHOY-DTPNARKL--NLNISTTKRHPRTIT 277
 QY 241 YSSTASDNT-----QSVAKGSTSKPANKSENLQSLKELYDPSAQ-RYSE-- 286
 DB 278 YLLDENNALPELDVLANLYLVKKNYVAHLQNLKNSYLLLGFRDENNNKIHSTPI 337
 QY 287 SDGLVF-----DPAKIISRPNGVALIPGHDIHYIPYS--KLSALEEKLIA 339
 DB 338 KTFILIGTFPHATIVFVNVNPKQIITSQDARIALDPVDDQKLRIGNPELYLKNKNQ 397
 QY 330 RMPVIGSTGTVSTNAKPNEVSSLSL--SSNSSLTTSKLSASADG--IFNKD-- 383
 DB 398 QVYKTIQYVGIKNTKNSGTLPELNLNLSNNYELLEELKALGLANDAYOKVVFAPDOK 457

```

2Y 384 -----IYEATATAYAHGSH-----FYIPRSNOI 409
Db 458 HNOFSTKLGITYLESISSEVYIKKDKNDLVNKIYFSONDKILNNVMAQIYISKDNEV 517
2Y 410 --GQP-----
Db 518 CLSQPLLLLVNGSNRYFTLLNNKPNGLSNRYKFKVLLIYADQNDLQSTDAVKNSTDL 577
2Y 413 -----TLPNNSLATPSP-----SLPDPGTSHKEHEEDGYGF-----D 445
Db 578 SHGFKTEPIGIISLKNNLSQVYIKRTDSDIRLSIPL-VQMDHYLVYDSDGSFTIKISDLD 636
2Y 446 ARIIIEDESGEFGVMSG-----DHNYFFPKDLTEQIKAK--OKHLEEV 488
Db 637 PNNLPDADOKIYIDONGNRNMLEFDVPVSLNKHKIHINITILLINKPINALAFQONNEHQV 696
2Y 489 KTSNGLDLSLSS-----HEODYGNAKEMD--LDKIE--EKIAGIMOYGRKRS 536
Db 697 YELANSQPOLSEFGNELLVNTYNTNPNITDVTNVTNLYKLLKSLMLMEYIKLY 756
2Y 537 IYVNEKRNALYPHGDHHDADPIDEKPYGIGHSHSNYELFPDEGVAKREGNKVTYGE 596
Db 757 IDNN-----GDVMSD-----PVSINNVTNEFEFNPDKALK--SNRIKFDG 798
2Y 597 L-----TNYNLIKSTFNQNFYTLANGQKRVSESPPELEKLGIMLVKLTDPGKVL 651
Db 799 LYFKDQNTVTKTSARKVMMNIT-----PLOIQASKIKINLSPIYV 840
2Y 652 EKVSQVGEVGVGNINAFELDQPLP-----GQFFKTYLASKQDPEVSYDQTFV 701
Db 841 -----INNTIPASADL-YLPVYSTDIDIPAKQVYIDVTIGNKD--NKNQDMFT 885
2Y 702 PSLAYKMASQITFPFHADDTYLAVN-----
Db 886 -SNLEEDDITKTGWYKIHANNLSPEVNYQIKSVKFRQKQVLAFAVNNNSDDHVLIDQK 944
2Y 729 -FOFAPKKTDLVRYTFDEFGNAYIENNYKGEIKLPKLNQGT-----774
Db 945 TGFSTPKASFPLVSY-----TASDVYDANKVKNSHVIVIN--NDGSLNCRKAKIYVNG 998
2Y 775 -----RTAGNKIPYTFMANAY-----LDNGSTIYVEPILEKENQD--KPSILPQFRNKA 824
Db 999 EHDLYSNRVYLVANGVNEDEPMLDKHNRITSFKKLITPENQNDYKSF-----1050
2Y 825 QENSKLDEKVEEPTSEKYEKE-KLSEGNSTSNSTLEEVPTVQOEKVAFAFASYGK 883
Db 1051 -NNOKITWTEKASVNEQITIDFKLYRKPSMLNKLINLQIND-----1094
2Y 884 LENVLFNMDGITELLPSEGVYKKNMADFTGAPQNGENKPSSENGKYSTGTVENQPTEN 943
Db 1095 --NDPLEDSDILEITFHDEK--DKTTHVYIGKINVDANNNKITLFESEVENTNFKIOPNKK 1150
2Y 944 KPADSILPEAPNEKPKVPENSTONGMLNBEANGVSGPMIDPALTEAPVAPVOEKL-----998
Db 1151 YIVDININATKKK--IQPANAISN--NSNKSITIYASSNPS-----KILLSFTN 1194
2Y 999 EKFTASYGIDSVIFNMDGITELLPSEGVYKKN 1033
Db 1195 ELFVNININQNPNTMLNPTASIDVELKSSQNLTKD 1229

```

RESULT 13
18447

hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)

Species: Plasmodium falciparum
Date: 15-Oct-1989 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

Accession: T18447
Lawson, D.; Bowman, S.; Bartell, B.

Submitted to the EMBL Data Library, August 1997

Accession: T18447

Status: preliminary; translated from GB/EMBL/DBD
Molecule type: DNA

A:Residues: 1-3724 <LAW>
A:Cross-references: EMBL:298547; NID:e1325376; PID:e1325379; PIDN:CAB11104.1
C:Genetics:
A:Introns: 307/1; 1545/2
A:Note: C0335c

Query Match 3.6%; Score 193; DB 2; Length 3724;
Best Local Similarity 16.7%; Pred. No. 0.52;
Matches 185; Conservative 185; Mismatches 381; Indels 358; Gaps 49;

```

QY 30 SOENKNNRNVSTVDSQ--SSQKSENIPTDQ-----VSQKEGIQ-----AEQIVIKITDQ 78
Db 446 NNEENIDNN--IYVENNKLKQDSYDLFSESEGNLILGVNEGEFEFNEFENIEKELQERK 503
QY 79 YVTSQDHHYNGVPPDALFSEELMKDPYQKADIVNEVGGIILVQDKYYIYL 138
Db 504 DEKNNDKTYINNNEQTD-----DLNRNINKTESIN-----535
QY 139 KDAADAVYRTDEINROKOEHVKNDEKYSNVAVARSGR-----YTTNDGYFNPADI 193
Db 536 NNDNNNNNNKKEKNTREHILNKESIKHIGSPSRDKKIKLYTNK-----NEDST 591
QY 134 IEDGNAYIVPHGHYHIIYPSDSLASELAALAKHLAGKNQPSQSYSTASDNNTOV 253
Db 532 FELKKELEIITNNKNVY--EEDIIGSNEDEYIHLKENLEKEDANETN-----638
QY 254 AKGSTSKPANKSENQSLKELDYDPSAQRYESDGLVFPDAKIISFTPNQVAPPHGHY 313
Db 639 -----NDKENNNKTLELKS--KNILENKRILEELK--RQRNNI-----676
QY 314 HEIPYSKLSALEEKIARVAPISGTSTVSTNAKPNEVYSSLSGSSNPSLTSKELSSA 373
Db 677 -----FKDEKYNLSGEVILINLIQNEKKINDI 705
QY 374 SDGYFNPRODYEETATAYIYRHGDHFIYIPKSNQIGOPTLPNNSLATPSPSLPINPTS 433
Db 706 QDGNISKQK--IIQSSS-----RTNDTFNI-----KDISLNDLE 738
QY 434 HEKHEEDGYDANILIEDSGFPMGSHDNHNYFFKDLTEQIKAKOKHLEEV-----488
Db 739 KEKRRKKSQHFIDNKVD-----KNEISENINKICDNNINNIYDESI 781
QY 489 -----KTSNGLDLSLSEHD-----YPNAKAKED--LDKILKEKTAGIMKOYGAR- 534
Db 782 NNIDSEINNIYDESINNIYDESINNIYDENINNIYDENINNIYDENINNIYDE-GIKKI 840
QY 535 -ESTIVNK-----EKNAIYPHGDHHDADPIDEKPYGIGHSHSNY 574
Db 841 CDDNILEKNKIKTTNDIYQVEENNESIEKNELM-----ISLKKDI-----NNTY 884
QY 575 ELFK-----PEEVAK--KEGN-----KYVTGE-----ELTN--599
Db 885 NMFKEVNDIFINKIRRESLKLIDKINIDNNODEYINDNFEENFIINHEKKEITNKLEDP 944
QY 600 -----VNL--LKNSTFNQNFYTLANGQKRVSESPPELEKLGIMLVKLTDPGK 649
Db 945 LEITQONEFIENDLIKKKYIYNDF--FNDADKAFYEAKLINDAKKKKQOEFPKFD--1000
QY 650 VLEKVSQVGEVGNINAFELDQPLPGQTFKTYIASKDYPEVSYDQTFVPTSLAYKM 709
Db 1001 -----ETFGLSQHKI--KYNNGEE--KHDKNNEEKNILYD-----1034
QY 710 ASQITFPFHAGDYLAVNPOFAPVKGIDALVRYTFDEHG--NAYLENNYKVEIKLPI 766
Db 1035 -----ENQYTSVLYSDHKIEDIDISHIQTICDENN-----I 1068
QY 767 PKLNQGTTRTGNKIPYTFMANAY--LDNGSTIYVEPILEKENQDPSILPQFRNKAQ 825
Db 1069 EQIIEENSK-KGVAISGTDEKNDAENKND-----MEKKNDMEKKNDIE--KKNDME 1118
QY 826 ENSKLDKVEEPTSEKYEKEKLSEGNSTSNSTLEEVPTVQOEKVAFAFASYGKLE 885

```



```

;Reference number: 219322
;Accession: T20774
;Status: preliminary; translated from GB/EMBL/DBJ
;Molecule type: DNA
;Residues: 1-4667 <N12>
;Cross-references: EMBL:281499; PDB:1CAB54224.1; GSPDB:GN00019; CESP:2K270.2d
;Experimental source: clone F11C3
;Kershaw, J.
submitted to the EMBL Data Library, November 1996
;Reference number: 220423
;Accession: T27818
;Status: preliminary; translated from GB/EMBL/DBJ
;Molecule type: DNA
;Residues: 1-4667 <N12>
;Cross-references: EMBL:282089; PDB:1CAB54513.1; GSPDB:GN00019; CESP:2K270.2d
;Experimental source: clone 2K270
;Genetics:
;Gene: CESP:2K270.2d
;Map position: 1
;Intons: 61/3; 112/1; 244/3; 328/2; 384/1; 406/3; 4302/3; 4343/3; 4384/1; 4422/3; 4461/3
Query Match 3.58; Score 190.5; DB 2; Length 4667;
Best Local Similarity 17.18; Pred. No. 0.99; Mismatches 448; Indels 443; Gaps 53;
Matches 219; Conservative 167;
57 DOVSQKEGIAEQIVIKITDQGYTSHGDHYHNGKVPYDALFSEELMDPNYQLDA 116
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2551 DEISEHNDLH---APIADANVYHGGDSRIEKHASPATLEKKOKKVASPKKEKDV 2606
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 117 DIVNEVGKGIYIKYDGKTYVYLKDAADANRYTDE-----INROKEVYKDNK 166
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 2607 PAEPKEKIRIJA-----RVATEPEAIVDPVYVKEKEKKGSKSLTP 2650
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 167 ---VNSVAVASGGRATTNDGYV--FNPAIDIED--GNAYIVPHGGHYIYKSDLSA 219
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 2651 TFGPFSKTKTKTKTGIVPEISEAYTGLDTRIHDLGTSF--EHGHPAYSFKQYIE 2708
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 220 S-----ELAAAKALAGKNM--QPSQSYSTASDNNTOYAK----- 255
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 2709 AIEPFAPEKERYLIAFRHGEDVDEYKPPQIATTLVLDGPLDEISRDTDHPPIA 2768
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 256 -----GSTSK-----PANKSENQSLKELIDSPSAQRYSESGLVFPDAKI 297
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 2769 DHAVNYHGGDSRIEKHASPATPEKKOKKVASPKKEKVDVPAPEPKE-----IRL 2820
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 298 ISRTPNGVAIHGGHYHPIPSKLSALEEKIARAVPISGTSVSTNAKPNVYVSLGSL 357
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 2821 IARVRY-----EPEAEIYDVVYVKEKEKKGSKSLPTFGFF 2858
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 358 SSNPSLTTT--KELSSASDGIYFNPKDIYETATAYIVRGGDHFIYIPKSNQIGOPTLP 415
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 2859 SKTTRTKTGTGYPETSEAYTGLDTRIHDLGTSF--EHGHPAYSFKQYI----- 2910
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 416 NNSLATPS--PSLPINPGTSHKEHEDC-----YGFANRITIAEDSGFVM 459
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 2911 -ETIETPAPEPKKRYRLIARRHGDEDEYKPPQIATTLVLDGPLDEISRDTD---L 2965
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 460 SH---GDNNHYFFKDLTDEQIKAAOKHLEVKTSHNGDLSLSSHEDYPPNAKEMDLD 516
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 2966 DHTPIADANVYHGGDSR-----TEKASPVAT-----LEKKOKKVASPKKEKVDV 3013
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 517 KRIEKTAGINKOYGVKRESIVN-----KEK-----NNIY 548
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 3014 AEPKKEIRLIARVTEPEAIVDPVYVKEKEKKGSKSLPTFGFSKTKTKTGTG 3073
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 549 PPGHNNADPID---EKKPVGIGSHSNYELFPEEGVA----- 584
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 3074 PETSEAYTGLDTRDHTDGTSTFEHGEHPAYSFKQYIATETPAPEPKKRYRLIARF 3133
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 585 KKEGNN-----YVTG--EELTANVVLKNSFTNNQNF----- 614
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 3134 RHGDEDEYKPPQIATTLVLDGPLDEISRDTDHTPIADANVYHGGDSRIEKHAS 3193
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

OY 615 --TLANGQRV-----SFSPELEKKLGINLVLLIT--PDGKYLEKVSQKVGEGVG 664
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3194 PVATLEKKOKKVASPKKEKVDVPAPEPKE--EIRLIARVTEPEAIVDPVYVKEKKG 3252
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 665 NIANFELDQYLLPQGTQK-----TIAKDQPEVS--YGS-----PFTVP 702
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3253 D-----SKSKLP--TGFFSKTKTKTKTGTPTSAYTGLDTRDHTDLEGTSTFEG 3304
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 703 TSLAYKMAQOTI-----FYPP-----HAGDTYLRNPPQFAVPRKGTAL 740
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3305 EHPAFEASSTITIEADPDSHEKVIKPSRFHMSRRRHGDEDEYKPPGGYI----- 3357
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 741 VRYFDEPHGNAYLENNYKVGELKPIPLNAGTTRTAGNKIPVTETMANAYLDNOSTIYE 800
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3358 --LSSEIYEGPY--DNMAPIGLEL-----TPILDHSQVYHNGSGKLEK 3398
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 801 VPILEKNQDTP-----SILPQFKRNKAQENSKLDEKVEEPTSEKY----- 843
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3399 RSSKEKIKTPSPERVPSEIKLVARIQPLEHEDSISEKLSPVKDRSRAPSFSMHRQ 3458
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 844 EKEK-----LSEIGNS--TGNSTLE--EVPYDVPQEKVAKFAESYGMKLENY 887
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3459 SKQRYEISPLTYEGNIDYGRALVEMENTPIEVASVPEIYHPKQGVSAQHES--SQPNR 3515
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 888 LFNNDGTIELYPSGEVYIKNMADFTGEADQG-----NGEN 923
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3516 FYRLIARRHGDEDELEKPDAYKFTVEYEGPLDIRPVVELADPISDSQVYHNGES 3575
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 924 -----KPSENGK-----VSGTGVNOPTENKPPADSLPEAP 953
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3576 WKAEKNIKPPPVANTKGLSPETSEVHEIRLLITVAVASSSEPEEVDVIPAESTVYP 3635
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 954 -----NEKPYKPNSTDNGMLNPEGNY-----GSDPMLDPALAEA 988
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3636 FRLAFLSIGKSPSPSEPEFCSPQYDGLDLSRSDLEVMYLLGVSVPYSPKFTPV 3695
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 989 PAVDPQEKLEKFTASY 1005
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3696 KKYDKADEERKKYVIY 3712
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

Search completed: September 26, 2001, 22:32:23
Job time: 307 sec.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

3M protein - protein search, using sw model

Run on: September 26, 2001, 21:37:01 ; Search time 57.46 Seconds
(without alignments)
1096 212 Million cell updates/sec

Title: us-09-471-255-2

Effect score: 5406
Gene: 1 MKSEKRTIAGSANTVLSLST.....IELRLPGEVYKRLNSPIA 1039

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: A: Geneseg.0601.*
1: /SIDSL/gcgdata/geneseg/genesegp/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseg/genesegp/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseg/genesegp/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseg/genesegp/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseg/genesegp/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseg/genesegp/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseg/genesegp/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseg/genesegp/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseg/genesegp/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseg/genesegp/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseg/genesegp/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseg/genesegp/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseg/genesegp/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseg/genesegp/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseg/genesegp/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseg/genesegp/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseg/genesegp/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseg/genesegp/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseg/genesegp/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseg/genesegp/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseg/genesegp/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseg/genesegp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|----------------------------|
| 1 | 5406 | 100.0 | 1039 | 21 | AA12715 Streptococcus pneu |
| 2 | 5315 | 98.3 | 1019 | 21 | AA12722 Streptococcus pneu |
| 3 | 5315 | 98.3 | 1019 | 21 | AA12751 Streptococcus pneu |
| 4 | 5297 | 98.0 | 1019 | 21 | AA12750 Streptococcus pneu |
| 5 | 5297 | 97.9 | 1019 | 21 | AA12748 Streptococcus pneu |
| 6 | 5291 | 97.9 | 1019 | 21 | AA12749 Streptococcus pneu |
| 7 | 5291 | 97.9 | 1019 | 21 | AA12752 Streptococcus pneu |
| 8 | 5291 | 97.9 | 1019 | 21 | AA12753 Streptococcus pneu |
| 9 | 4276.5 | 79.1 | 840 | 21 | AA12721 Streptococcus pneu |
| 10 | 4090 | 75.7 | 780 | 21 | AA12744 Streptococcus pneu |
| 11 | 2953 | 54.6 | 568 | 21 | AA12731 Streptococcus pneu |

| | | | | | |
|----|--------|------|------|----|---------|
| 12 | 2953 | 54.6 | 1057 | 21 | AA12725 |
| 13 | 2746 | 50.8 | 528 | 21 | AA12719 |
| 14 | 2650 | 49.0 | 509 | 21 | AA12724 |
| 15 | 2559 | 47.3 | 489 | 21 | AA12723 |
| 16 | 2494.5 | 46.1 | 485 | 21 | AA12738 |
| 17 | 2492 | 46.1 | 484 | 21 | AA12767 |
| 18 | 2492 | 46.1 | 484 | 21 | AA12708 |
| 19 | 2485 | 46.0 | 484 | 21 | AA12718 |
| 20 | 2338 | 43.2 | 447 | 19 | AA12728 |
| 21 | 1728 | 32.0 | 329 | 21 | AA12732 |
| 22 | 1307 | 24.2 | 819 | 21 | AA12716 |
| 23 | 1295.5 | 24.0 | 819 | 21 | AA12768 |
| 24 | 1286 | 23.8 | 819 | 21 | AA12769 |
| 25 | 1286 | 23.8 | 819 | 21 | AA12720 |
| 26 | 1282.5 | 23.7 | 826 | 21 | AA12739 |
| 27 | 1280.5 | 23.7 | 838 | 21 | AA12766 |
| 28 | 1277.5 | 23.6 | 827 | 21 | AA12762 |
| 29 | 1266.5 | 23.4 | 821 | 21 | AA12727 |
| 30 | 1266.5 | 23.4 | 821 | 21 | AA12766 |
| 31 | 1264.5 | 23.4 | 690 | 21 | AA12745 |
| 32 | 1263.5 | 23.4 | 811 | 21 | AA12763 |
| 33 | 1263.5 | 23.4 | 834 | 21 | AA12739 |
| 34 | 1249.5 | 23.1 | 807 | 21 | AA12765 |
| 35 | 1247 | 23.1 | 796 | 19 | AA12760 |
| 36 | 1243.5 | 23.0 | 811 | 21 | AA12760 |
| 37 | 1243.5 | 23.0 | 811 | 21 | AA12762 |
| 38 | 1240.5 | 22.9 | 816 | 21 | AA12756 |
| 39 | 1239.5 | 22.9 | 816 | 21 | AA12756 |
| 40 | 1238.5 | 22.9 | 819 | 21 | AA12740 |
| 41 | 1238.5 | 22.9 | 819 | 21 | AA12754 |
| 42 | 1237.5 | 22.9 | 805 | 21 | AA12764 |
| 43 | 1236.5 | 22.9 | 811 | 21 | AA12761 |
| 44 | 1233 | 22.8 | 820 | 21 | AA12755 |
| 45 | 1231.5 | 22.8 | 816 | 21 | AA12758 |

ALIGNMENTS

RESULT 1
ID AA12715 standard; Protein; 1039 AA.
XX
AC AA12715;
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae BVH-3 protein antigen SEQ ID NO:2.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-26; antigen; vaccine; prophylaxis; therapy; infection; diagnosis; meningitis; bacteremia; otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
(BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Chailand N;
XX
DR MPI: 2000-452397/39.
XX
PT N-PSDB; AAA65730.
XX
OT Streptococcal antigens useful for vaccinating against e.g. meningitis, otitis media, bacteremia and/or pneumonia

35 Claim 18; Fig 2; 106pp; English.

36 The present invention describes nucleic acids (I) encoding protein

37 antigens (II) from Streptococcus pneumoniae. The protein antigens

38 have bactericidal activity. The nucleic acids, encoding the protein

39 antigens, may be used for the recombinant production of the proteins

40 they encode. The protein antigens may then be used as vaccines for the

41 prevention and treatment of Streptococcal infections in mammals

42 (especially humans) which result in, e.g. meningitis, otitis media,

43 bacteraemia and/or pneumonia. The present sequence represents the

44 S. pneumoniae BVH-3 protein antigen.

45 Sequence 1039 AA;

Query Match 100.0%; Score 5406; DB 21; Length 1039;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1039; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKFSKRYIAAGSAVIVSLCAVALNHRQENKDNMRVSYVDSOSSOKSENLTPOVDS 60

1 mkfskryiaagsavivslcavalnhrsqenkdnmrvsyvdsqsgksenltpdqvs 60

61 QKEGIAQDIYKIDQGVYSHGSHYHYNKVRYDALFSEELAKPBNYQKADIVN 120

61 qkegiaqdiiykidqgvysghshyhyngkvyrydaliseellmkpbnypqkadiyn 120

121 EYKGGYIIVDKYVYVYLDAAHADNVRTKDEINRQKEHVKNKNSNVAVARSQRY 180

121 eykgyiivdkyvyvyladaahadvrtkdeinrqkehvknsnvavarsqry 180

181 TNDGIVYRPAIDIEDTGNAVYVPHGSHYHYTPKSDLSASEIAAKAHLAGKMPQSLS 240

181 tndgivyrvpaiedetgnavyvpghshyhytpksdlsaseiaaahlagkmpqs 240

241 YSSTASDNNTQSVAGSGSKPANKSENLOSLKELYDPSAQRESGIVYDPDKITSR 300

241 ysstasdnntqsvagsgskpanksenloslkeltydpsaqresgivydpdkit 300

301 TPNGVAVPHGSHYHYTPYKSLALEEKIARWPIGSGTSTVSTNAKPNEVVSLSLSLSN 360

301 tpngvavphgshyhytpyskaleekiarmwpihgsgtstvstnkpnevvs 360

361 PSSLTSELSASAGYIFNPKDIYEETATAYVARGHPIYKPSKNOIGOPTLPNSLA 420

361 pssltselsasagytfnpkdiyeetataiyvarghpiykpknogoptlpnsla 420

421 TPSPSLPINPGTSHKHEHDGYPGPNARLIIAEDSGFYMSGDNHNYEKKDITEQIKA 480

421 tpspslpinpgtshkhehdgypgnarliliaedsgfymsgdnhnyeekdit 480

481 AOKHLEVTSHNGIDSLSSHEODYPGNAKEMKDKRIEIKIAGIMKQVYKRESIYVN 540

481 aokhlevtshngidslssheodypgnakemkdkrieikiagimkvkyresiy 540

541 KEKNAITTPGDDHHADIDDEKPYGIGSHSNTLEKPEEGVAKKEGNKYTGEBELTNV 600

541 keknaitypgddhhadiddekpygigshsntlekpegvakkegnkytgebel 600

601 VNLKNSFNNOFTLIANGOKRVSPFPELEKLGIMVAKLITPOKLYEYSGKATG 660

601 vnlknsfnnoftliangokrvspfpeleklgimvacklitpokyeysgkatg 660

661 EGVGNIANFELDQPTLPQGTQKTYTASKDYPEVSIDGTFYPTSLAYMASQITFYFPA 720

661 egvgnianfeldqptlpqgtqktytasdkdypevsidgtfypstlaymasq 720

721 GDTYLRVAPQAVRKGTALVAVDEFGNAYLENNYVGGIKLPIPLNOGTRTAGNK 780

721 gdtylrvapqavrkgtalvavdefgnaylennvvggiklpiplnogtrtag 780

781 IPTFMANAYLNDQSTYVVEPILEKENQTKPSILPOFKRNKAQENSKDEKYEERTS 840

781 iptfmanaylndqstyvvepilekenqtkpsilpofkrnkaqenskdekye 840

Db 781 iptfmanaylndqstyvvepilekenqtkpsilpofkrnkaqenskdekye 840

QY 841 EKVEKEKLETSNGSTNSSTLEEVPTVDPQEVKAFASYSQKLENVLFNNDGTIELYLP 900

841 ekvekekletsgstnsstleevptvdpqevkafasysqklevnlfnndgtiel 900

Db 841 ekvekekletsgstnsstleevptvdpqevkafasysqklevnlfnndgtiel 900

QY 901 SGEVYKKNMADFTGAPQNGENKPSKNGKYSTGVENQPTENKPADSLPEAPNEKPYKP 960

901 sgevyykknmadftgapngenkpskngkystgvnqptenkpadslpeapnek 960

Db 901 sgevyykknmadftgapngenkpskngkystgvnqptenkpadslpeapnek 960

QY 961 ENSDNGMLNPEGNGSPMDPALAEAPVDPQEKLEKFTASGIGLDSYIFNMOSTI 1020

961 ensdngmlnpegngspmdpalaeapvdpqeklekftasgigldsyifnmosti 1020

Db 961 ensdngmlnpegngspmdpalaeapvdpqeklekftasgigldsyifnmosti 1020

QY 1021 ELRLPSGEVYKKNLSDFIA 1039

1021 elrlpsgevykknlsdfia 1039

Db 1021 elrlpsgevykknlsdfia 1039

RESULT 2

AAB12722

ID AAB12722 standard; Protein: 1019 AA.

XX

AC AAB12722;

XX

DT 21-NOV-2000 (first entry)

XX

DE Streptococcus pneumoniae BVH-3M protein antigen SEQ ID NO:55.

XX

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

KW otitis media; pneumonia; immunisation; bactericidal.

OS Streptococcus pneumoniae.

XX

PN MO200039299-A2.

XX

PD 06-JUL-2000.

XX

PF 20-DEC-1999; 99MO-CA01218.

XX

PR 23-DEC-1998; 98US-0113800.

XX

PA (BIOC-) BIOCHEM PHARMA INC.

XX

PI Hamel J, Brodeur BR, Plneau I, Martin D, Rioux C, Charland N;

XX

DR WPI: 2000-452397/39.

XX

XX Streptococcal antigens useful for vaccinating against e.g. meningitis,

PT otitis media, bacteraemia and/or pneumonia -

XX

PS Claim 18; Fig 20; 106pp; English.

XX

CC The present invention describes nucleic acids (I) encoding protein

CC antigens (II) from Streptococcus pneumoniae. The protein antigens

CC have bactericidal activity. The nucleic acids, encoding the protein

CC antigens, may be used for the recombinant production of the proteins

CC they encode. The protein antigens may then be used as vaccines for the

CC prevention and treatment of Streptococcal infections in mammals

CC (especially humans) which result in, e.g. meningitis, otitis media,

CC bacteraemia and/or pneumonia. The present sequence represents the

CC S. pneumoniae BVH-3M protein antigen.

XX

Sequence 1019 AA;

Query Match 98.3%; Score 5315; DB 21; Length 1019;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 CAYALNHRQENKDNMRVSYVDSOSSOKSENLTPOVDSQKEGIAQDIYKIDQGVY 80

21 cayalnhrsqenkdnmrvsyvdsqsgksenltpvdsqkegiaqdiiykidqgv 80

```

> 1 cayalnhrgenkdnnrvsyvdsqsgskseanltpdqvsqkqegiaeqivlkltdggy 60
> 81 TSHGCHYHYHNKRVYDALFSEELMKDPNYOLKADYNEKGGYIIKVDGKYVYVYKLD 140
> 61 tshgchhyhnygkvpdydalfeellmkdpnyqldkdadlvnevggylkvdgkyvyyikld 120
> 141 AAHANVRFKDKDINKOKHRYDNKRVNSNVAVARSGRYTNDGYVFNPAIIEDTGNA 200
> 121 aahadvrftkdelnqkghvkdnekvnsnavarsqgytndgyvfnpadlledtgna 180
> 201 YIVPHGCHYHYTPKSDLSASELAALKAHLAGKNMOPSOLESYSTASDNNTQGVANGSTSK 260
> 181 yivphgchhyhytpksdlsaseelaakahlagknmpsqysystasdnntgsvangstsk 240
> 241 pankeenlqslldkelydpsaqryesdglvfdpakllartpnyvalpbgdhyhflpyak 300
> 321 LSALTEKIRAMVPISGTSVSTNAKPNEVSSIGLSNPSLITTSKELSSASDGYTFN 380
> 301 lsaleeklarmvpisgtsvstnakpnevssigslnpslittskelssasdgtyfn 360
> 381 PRDIYEETATAYVRHGDHFHYTPKSNQIGOPTLPNNSLATSPSLPINPGTSHKHEED 440
> 361 prdiveetatayivrghdghfhytpkngdqgprlpnslatpsslpinpgtshkheed 420
> 441 GTFEDANRIIADESGTFVNSHGDHNYFFPKDITREQIQAQAQHLIEVYTSNGLDLSS 500
> 421 gtfedanriiaeesgtfvnsghdghnyffkdliteeqilaahqhlievtsngldlss 480
> 501 HBDDYGNKNEKMDLTKTEEKIAGIMKQYGRRESIVYNNKKNALITYHGHNHADPID 560
> 481 hbddypgnkemdldtkleeklagimkqygrresivnnkknalityghdnhadpid 540
> 561 EHRPVGIGSHSHYELFEKFEDEGVAKKEGKRVYTGELITVNVMLKKNSTNNONFTLANQ 620
> 541 ehrpvigshshyelfekfedegvakkegkrvytgelitvnmvmlkknstnnonftlanq 600
> 621 KRYSFEPPELEKKGILGNLVKILITPDGKYLEKVSQKTVGEGVGNANFELDOPYIPGCT 680
> 601 krysfeppelekkgilgnlvkilitpdgkylekvsqkvegvggnanfelqpyipgct 660
> 681 FRYTASKRYPEVSYDGTPTPSLAUKKASOTIFYPHAGPTYLKVNPOFAPVKGTDAL 740
> 661 frytasrkrypevsydgtpptpslaulkkasotifypahagptylkvnpofoapvkgtdal 720
> 741 VRYFDEFGHNAYLENNYKVGELIKLPIPKINOGTTRPAGNKIPVTFEAMNAYLDNOSTYIYE 800
> 721 vryfdeghgnaylenmykvgekliprpklingttrtagnkipvtfamanyldngstyie 780
> 801 VPILKEKNTDKPSIIPORRNKAQDNKULDEKVEEPKTSERVEKESKLSGTNSNSTL 860
> 781 vpilkeengtdkpsilppgrnkagdnkuldekveepktservkeklsklsgtnsnstl 840
> 861 EEPVTPDPOEVAKFAEYKGLKLENVLFNMDGETIELYPSGEVIKKKNADTGEAPQDN 920
> 841 eepvtpdpovevakfaeykglklenvlfnmogetlelypsgevikknaditgeapqdn 900
> 921 GENKSENGKRVSTGTVENOPTENKRPADSLPEAPNEKRVPRKSTNDGMLNPGNNGVSDPM 980
> 901 genkspengkvtstgtenpntenkpadslepapnekrvprkstndgmlnpgnngvsgdpm 960
> 981 LDPALFEAPVDPVOEKEKFTASVGLSDSYTFNMKGSTIELRPSGEVYIKKNLSDFIA 1039
> 961 ldpaleeapavdpvoekektasvlgldsytfnmkgstielrpsgevikknlsdfia 1019

```

RESULT 3
 ID AAB12751 standard; Protein; 1019 AA.
 AC AAB12751;
 XX

```

> 21-NOV-2000 (first entry)
> DE Streptococcus pneumoniae strain SP64 BVH-3 protein antigen.
> XX
> DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
> KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
> KW otitis media; pneumonia; immunisation; bactericidal.
> XX
> OS Streptococcus pneumoniae.
> XX
> PN MO200039299-A2.
> PD 06-JUL-2000.
> XX
> PF 20-DEC-1999; 99WO-CA01218.
> XX
> PR 23-DEC-1998; 98US-0113800.
> XX
> PA (BIOC-) BIOCHEM PHARMA INC.
> PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
> DR WPI; 2000-453397/39.
> XX
> PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
> PT otitis media, bacteraemia and/or pneumonia
> XX
> PS Disclosure: Flg 11; 106pp; English.
> XX
> CC The present invention describes nucleic acids (I) encoding protein
> CC antigens (II) from Streptococcus pneumoniae. The protein antigens
> CC have bactericidal activity. The nucleic acids, encoding the protein
> CC antigens, may be used for the recombinant production of the proteins
> CC they encode. The protein antigens may then be used as vaccines for the
> CC (especially humans) which result in, e.g. meningitis, otitis media,
> CC bacteraemia and/or pneumonia. The present sequence represents a
> CC S. pneumoniae BVH-3 protein antigen, from the present invention.
> XX
> SQ Sequence 1019 AA;

```

Query Match 98.3%; Score 5315; DB 21; Length 1019;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

> QY 21 CAYALNHRSGENKDNRRVSYVDSQSOSKSENLPDOVSQREGIOAEQIVYIKITDQGY 80
> 1 cayalnhrgenkdnnrvsyvdsqsgskseanltpdqvsqkqegiaeqivlkltdggy 60
> QY 81 TSHGCHYHYHNKRVYDALFSEELMKDPNYOLKADYNEKGGYIIKVDGKYVYVYKLD 140
> 61 tshgchhyhnygkvpdydalfeellmkdpnyqldkdadlvnevggylkvdgkyvyyikld 120
> QY 141 AAHANVRFKDKDINKOKHRYDNKRVNSNVAVARSGRYTNDGYVFNPAIIEDTGNA 200
> 121 aahadvrftkdelnqkghvkdnekvnsnavarsqgytndgyvfnpadlledtgna 180
> QY 201 YIVPHGCHYHYTPKSDLSASELAALKAHLAGKNMOPSOLESYSTASDNNTQGVANGSTSK 260
> 181 yivphgchhyhytpksdlsaseelaakahlagknmpsqysystasdnntgsvangstsk 240
> QY 241 PANKEENLQSLLDKELYDPSAQRYESDGLVFDPAKLLARTPNYVALPBGDHYHFLPYAK 300
> 321 LSALTEKIRAMVPISGTSVSTNAKPNEVSSIGLSNPSLITTSKELSSASDGYTFN 380
> 301 lsaleeklarmvpisgtsvstnakpnevssigslnpslittskelssasdgtyfn 360
> QY 381 PRDIYEETATAYVRHGDHFHYTPKSNQIGOPTLPNNSLATSPSLPINPGTSHKHEED 440
> 361 prdiveetatayivrghdghfhytpkngdqgprlpnslatpsslpinpgtshkheed 420

```

```

1Y 441 GIGFDANRIIAEDSGFVMSHGDHNYFFKDLTEEOIKAAQKHLEEVTSNGLDLSLS 500
1Y 421 gygfdanriiaedsgfvmshgdnhyffkdlteeqikaaqkhleevtsnngldsls 480
1Y 501 HEODYGNKEMKDLCKIEEKIAGIMKGYKRESTIVNKEKNALIIYHGHHAADPTD 560
1Y 481 heodygnkemekdlckieeklagimkgykreestivnkeknaliiyphghhadrpid 540
1Y 561 EHKPVGIGHSHSNYELFKPEDEGVAKEKNGKVTGEEELTVNVLKKNSTNNQNTLANCO 620
1Y 541 ehkpvigishsnayelfkpeegvakengkytgeeltvynvllknstnnqntlanq 600
1Y 621 KRVSEFPPELEKKGIMLVKLTIPDGKYLEKVSQKVEGEGVGNANFELDQPYLPQGT 680
1Y 601 krvsfeppelekkigimlvkltpdgkylekvsqkvegegvgnanfelddpylpqgt 660
1Y 681 EFKTIASKDYPEVSYDGTFTVPSLAKMASQTIFFPBHAGDTYLRVNPQFAPVKGTDAL 740
1Y 661 fkytiaskdypevsydgftftvpslaksasqtlffpbhagdtlyrvnpqfapvkgtdal 720
1Y 741 VRVDFEFGHNAVLENNTKYGEIKLPIPKINOGTTRTAGNKIPTFMANAYLDNOSTYIYE 800
1Y 721 vrdfefghnavlenntykgeiklppiknognotttagnktpfmanayldngstyive 780
1Y 801 VPILEKNOTDKPSIIPQFRKNAKQENSLDERKVEEPTSEKVEKEKLESTGNSTNSNTL 860
1Y 781 vpileknotdkpsiiipqfrnkakqenslderkevpeptsekevkeklesetgnstnsntl 840
1Y 861 EEPYTPDPOEVAKFAEYSGMKLENVLEFNDGTIELYPSSEVYIKKNADTGEAPQGN 920
1Y 841 eepytvdpoevakfaesygmkleenvlefnndgtielypssevyikknadtgeapqgn 900
1Y 921 GENKSENGKVTGTEVGNPTENKRPADSLPEAPNEKPVPEKSTNDGMNPGNNGSDPM 980
1Y 901 genksengkvtgtvgnptenkrrpadslpeapnekpvpekestndgmnpngngsdpm 960
1Y 981 LDPALAEADVPDVEKLEKFTASTYSGLDVYIFNDGTIELRLPSGEVYIKKNLSDPTA 1039
1Y 961 ldpalaeadvpdveklekftastystysgldvyifndgtielerlpsgevyikknlsdpta 1019

```

```

RESULT 4
AB12750
D AAB12750 standard; Protein; 1019 AA.
C AAB12750;
T 21-NOV-2000 (first entry)
E Streptococcus pneumoniae strain JNR7/87 BVH-3 protein antigen.
S Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
W prophylaxis; therapy; infection; diagnosis; meningitis; bacteremia;
V otitis media; pneumonia; immunisation; bactericidal.
S Streptococcus pneumoniae.
X MO200039289-A2.
X 06-JUL-2000.
X 20-DEC-1999; 99WO-CA01218
X 23-DEC-1998; 98US-0113800
X (BIOC-) BIOCHEM PHARMA INC.
X Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
R WPI; 2000-452397/39.
S Streptococcal antigens useful for vaccinating against e.g. meningitis.

```

```

PT otitis media, bacteremia and/or pneumonia
XX Dislosure; Fig 11; 106pp; English.
PS The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteremia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-3 protein antigen, from the present invention.
XX Sequence 1019 AA;
SQ
Query Match 98.0%; Score 5297; DB 21; Length 1019;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1016; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
1Y 21 CAYALNORSOENKDNRRVSYVDGSSQSKSENLPDQVSOKEGTOAQOYIKITDQGY 80
1Y 1 cayalnqrsqenkdnnrvsyvdgssqsksenlpdqvsqkegtoaqoyivakitdggyv 60
1Y 81 TSHGDHYHYNGKRYVDALFSEELMKDQNTOLKADIVNEKGGYIIRKVGKYYVYKLD 140
1Y 61 tshgdyhyhyngkyvpydalfsaelmkdqnpylkdadivnekggyiirkvgkyyvylkd 120
1Y 141 AAHADNVRTKDEINOKQBHVKNKENVSNVAVASQGRYTTNDGYVFNPAIDTIEDTNA 200
1Y 121 aahadnvtkdeinokqbhvknkenvsnvavarsqgryttndgyvfnpaditiedtnga 180
1Y 201 YIVPBGHYHYTPKSDLSASELAAKALAGKNMPSQLASTSRTSDNNNTGVAKGTSK 260
1Y 181 yivpbghyhytpksdlsaseelaakalagknmpsqlyststsdnnntgvakgtsk 240
1Y 261 PAMKSENQSLKELDYDSPAQRSESDGLVFDPKIIISRPNGVAIAPHGHYHPIPSK 320
1Y 241 pamksenqslkeldydspaqrsestdglvfdpkiiisrpngvaihghyhipysk 300
1Y 321 LSALEEKTAARVPIISGTSVSTNAKPNVSVSSGSSSNPSLTSKELSSASDGYTFN 380
1Y 301 lsaleektarvpiisgtsvstnakpnvsvssgsssnpsltskelssasdgtyfn 360
1Y 381 PKDIVEEATAYIYRHGDHRYIKPSNOIGOPTLPNLSAPPSGLPNPCTSEKHEED 440
1Y 361 pkdiveeatayiyrhgdhryipksnigoptlpnlsappsglpnpctsekehned 420
1Y 441 GYGFDANRIIAEDSGFVMSHGDHNYFFKDLTEEOIKAAQKHLEEVTSNGLDLSLS 500
1Y 421 gygfdanriiaedsgfvmshgdnhyffkdlteeqikaaqkhleevtsnngldsls 480
1Y 501 HEODYGNKEMKDLCKIEEKIAGIMKGYKRESTIVNKEKNALIIYHGHHAADPTD 560
1Y 481 heodygnkemekdlckieeklagimkgykreestivnkeknaliiyphghhadrpid 540
1Y 561 EHKPVGIGHSHSNYELFKPEDEGVAKEKNGKVTGEEELTVNVLKKNSTNNQNTLANCO 620
1Y 541 ehkpvigishsnayelfkpeegvakengkytgeeltvynvllknstnnqntlanq 600
1Y 621 KRVSEFPPELEKKGIMLVKLTIPDGKYLEKVSQKVEGEGVGNANFELDQPYLPQGT 680
1Y 601 krvsfeppelekkigimlvkltpdgkylekvsqkvegegvgnanfelddpylpqgt 660
1Y 681 EFKTIASKDYPEVSYDGTFTVPSLAKMASQTIFFPBHAGDTYLRVNPQFAPVKGTDAL 740
1Y 661 fkytiaskdypevsydgftftvpslaksasqtlffpbhagdtlyrvnpqfapvkgtdal 720
1Y 741 VRVDFEFGHNAVLENNTKYGEIKLPIPKINOGTTRTAGNKIPTFMANAYLDNOSTYIYE 800
1Y 721 vrdfefghnavlenntykgeiklppiknognotttagnktpfmanayldngstyive 780

```

2Y 801 VPILEKENTDKPSIIIPQFRNKAQENSKLDEKVEEPTSEKYEKEKLESETGNTSTNLT 860
 DB 781 VPILEKENTDKPSIIIPQFRNKAQENSKLDEKVEEPTSEKYEKEKLESETGNTSTNLT 840
 2Y 861 EEPYTPVPOEKYAKFAESYGMKLENTLFNMDGTIELYIPSGEVYIKKMADEFGAEPQGN 920
 DB 841 EEPYTPVPOEKYAKFAESYGMKLENTLFNMDGTIELYIPSGEVYIKKMADEFGAEPQGN 900
 2Y 921 GEKPPSENGKVSIGTYENQPTENKRPADSLPEANNEKVPKRENTDGMKNPEGNVSDPM 980
 DB 901 GEKPPSENGKVSIGTYENQPTENKRPADSLPEANNEKVPKRENTDGMKNPEGNVSDPM 960
 2Y 981 IDPALAEAPVADVOKLEKFTASYGGLDSYVFNMDGTIELRLPSGEVYIKKMLSDPIA 1039
 DB 961 IDPALAEAPVADVOKLEKFTASYGGLDSYVFNMDGTIELRLPSGEVYIKKMLSDPIA 1019

RESULT 5
 AAB12748
 ID AAB12748 standard; Protein; 1019 AA.
 AC AAB12748;
 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae strain W02 BVH-3 protein antigen.
 KM Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 NM Propylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 XX Otitis media; pneumonia; immunisation; bactericidal.
 OS Streptococcus pneumoniae.
 XX MO200039299-A2.
 PD 06-JUL-2000.
 XX 20-DEC-1999; 99MC-CA01218.
 PF 23-DEC-1998; 98GS-0113800.
 PR (BIOC-) BIOCHEM PHARMA INC.
 PT Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX WPI; 2000-452397/39.
 DR Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT Otitis media, bacteraemia and/or pneumonia -
 PS Disclosure; Fig 11; 106pp; English.
 XX

The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the proteins
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-3 protein antigen, from the present invention.
 CC
 SQ Sequence 1019 AA;

Query Match 97.9% Score 5291; DB 21; Length 1019;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 1015; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 21 CAVALNORHROENKNNRVSVDGSSQSKSENTPDOVSOKEGIOAEOIVIKITDGYV 80
 DB 1 CAVALNORHROENKNNRVSVDGSSQSKSENTPDOVSOKEGIOAEOIVIKITDGYV 60

QY 81 TSHGDHYHYNGKVPYDALFSEELMKDNPYOLKADIVNEVAGGIIIVDGATYYTLEO 140
 DB 61 TSHGDHYHYNGKVPYDALFSEELMKDNPYOLKADIVNEVAGGIIIVDGATYYTLEO 120
 QY 141 AAHADNVRKDEINRQKQHVADNEKVNNSNVAASQGVTTNDGVYFPAIDIEETGNA 200
 DB 121 AAHADNVRKDEINRQKQHVADNEKVNNSNVAASQGVTTNDGVYFPAIDIEETGNA 180
 QY 201 YIYPHGHHYIPKSDLSASELAARAKLAGKKNQPSQSLSYSTADNDTQSAKGSSTK 260
 DB 181 YIYPHGHHYIPKSDLSASELAARAKLAGKKNQPSQSLSYSTADNDTQSAKGSSTK 240
 QY 261 PANKSENLQSLKELTSDPSAORYSESDGLVEDPAKISRTPNQVALIPGHDIYHFIPIYK 320
 DB 241 PANKSENLQSLKELTSDPSAORYSESDGLVEDPAKISRTPNQVALIPGHDIYHFIPIYK 300
 QY 321 LSALEEKIARMPISGTGSTVSTNAKPNEVSSLSLSNPSSTLSKELSSASDQYIN 380
 DB 301 LSALEEKIARMPISGTGSTVSTNAKPNEVSSLSLSNPSSTLSKELSSASDQYIN 360
 QY 381 PKDIVETATAYIVRHGDHPHYIPKSNQIGOPTLPNNSLATPSPSLPIPGTSHKHEED 440
 DB 361 PKDIVETATAYIVRHGDHPHYIPKSNQIGOPTLPNNSLATPSPSLPIPGTSHKHEED 420
 QY 441 GYGFANRITIAEDSEGFVASHGDHNYFFKDLTEQIKAAOKHLEEVKTSIHNGDLSLS 500
 DB 421 GYGFANRITIAEDSEGFVASHGDHNYFFKDLTEQIKAAOKHLEEVKTSIHNGDLSLS 480
 QY 501 HEODYRGNKEMKDLKIEKTAGIMKOYKVRRESIVYKKEKNAIITPHGDHNDPID 560
 DB 481 HEODYRGNKEMKDLKIEKTAGIMKOYKVRRESIVYKKEKNAIITPHGDHNDPID 540
 QY 561 EHKPVGIGHSNHYELFKPEGVAKKGNKVYTGELTNVNLKSTFNQNTLANGO 620
 DB 541 EHKPVGIGHSNHYELFKPEGVAKKGNKVYTGELTNVNLKSTFNQNTLANGO 600
 QY 621 KRVSFSPPELEKRIKGINMLVKITPDGKYLEKVSQKVEGVGNLANEELDOPYLPQGT 680
 DB 601 KRVSFSPPELEKRIKGINMLVKITPDGKYLEKVSQKVEGVGNLANEELDOPYLPQGT 660
 QY 681 FKYTIAKDYPEVSYDGTFTVPTSLAYKMASTQTFYFPAAGDTYLRVNPQFAPKGTDL 740
 DB 661 FKYTIAKDYPEVSYDGTFTVPTSLAYKMASTQTFYFPAAGDTYLRVNPQFAPKGTDL 720
 QY 741 VRYFDEPHGNAYLENNYKVEIKLPIPKLNOGTTFRAGNKIPPTFMANAYLDNOSRYIVE 800
 DB 721 VRYFDEPHGNAYLENNYKVEIKLPIPKLNOGTTFRAGNKIPPTFMANAYLDNOSRYIVE 780
 QY 801 VPILEKENTDKPSIIIPQFRNKAQENSKLDEKVEEPTSEKYEKEKLESETGNTSTNLT 860
 DB 781 VPILEKENTDKPSIIIPQFRNKAQENSKLDEKVEEPTSEKYEKEKLESETGNTSTNLT 840
 QY 861 EEPYTPVPOEKYAKFAESYGMKLENTLFNMDGTIELYIPSGEVYIKKMADEFGAEPQGN 920
 DB 841 EEPYTPVPOEKYAKFAESYGMKLENTLFNMDGTIELYIPSGEVYIKKMADEFGAEPQGN 900
 QY 921 GEKPPSENGKVSIGTYENQPTENKRPADSLPEANNEKVPKRENTDGMKNPEGNVSDPM 980
 DB 901 GEKPPSENGKVSIGTYENQPTENKRPADSLPEANNEKVPKRENTDGMKNPEGNVSDPM 960
 QY 981 IDPALAEAPVADVOKLEKFTASYGGLDSYVFNMDGTIELRLPSGEVYIKKMLSDPIA 1039
 DB 961 IDPALAEAPVADVOKLEKFTASYGGLDSYVFNMDGTIELRLPSGEVYIKKMLSDPIA 1019

RESULT 6
 AAB12749
 ID AAB12749 standard; Protein; 1019 AA.
 AC AAB12749;
 21-NOV-2000 (first entry)

Streptococcus pneumoniae strain RX1 BVH-3 protein antigen.
 Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 prophylaxis; therapy; infection; diagnosis; meningitis; bacteremia;
 otitis media; pneumonia; immunisation; bactericidal.
 Streptococcus pneumoniae.
 MO200039299-A2.
 06-JUL-2000.
 20-DEC-1999; 99MO-CA01218.
 23-DEC-1998; 98US-0113800.
 (BIOC-) BIOCHEM PHARMA INC.
 Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 WPI; 2000-452397/39.
 Streptococcal antigens useful for vaccinating against e.g. meningitis,
 otitis media, bacteremia and/or pneumonia.
 Disclosure; Fig 11; 106pp; English.
 The present invention describes nucleic acids (I) encoding protein
 antigens (II) from Streptococcus pneumoniae. The protein antigens
 have bactericidal activity. The nucleic acids, encoding the proteins
 antigens, may be used for the recombinant production of the proteins
 they encode. The protein antigens may then be used as vaccines for the
 prevention and treatment of Streptococcal infections in mammals
 (especially humans) which result in, e.g. meningitis, otitis media,
 bacteremia and/or pneumonia. The present sequence represents a
 S. pneumoniae BVH-3 protein antigen, from the present invention.

Query Match 97.9%; Score 5291; DB 21; Length 1019;

Best Local Similarity 99.5%; Pred. No. 0; Mismatches 2; Indels 0; Gaps 0;

Matches 1014; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

21 CAYALNQRSEKNNKNNRVSYVDSQSSQSENLTPDVOSEKIGIADQIYKIDQGYV 80
 1 cayalnqrseknknrvsyvdsqssqsenltprdvsgsqegiaeqivikltdqgyv 60
 81 TSHGDHYHYNGKVPYDALFSEBLMKDPNTOLKADIVNEVKGYYIIRVDGKYYVILKD 140
 61 tshgdhyhyngkvydal fseblmkdpntolkadivnevkgyyiirkvdgkyyvilkd 120
 141 AAHADNVTKDEINNOKEHYKDNKENVNVAVARSGRYTNGCYVNPADIIEDTGN 200
 121 aahadnvtkdeinnokeh ykdnkenvnva varsgrytngcyvnpadiiedtgn 180
 201 YVPHGSHYHYTPKSDLSASELAAKAHIAKRNKQPSOLSTYSFASDNNQSVAKGSSTK 260
 181 ylvphgshyhytpksdlsase laa kahnkqpsolstysf asdnnqsvakgsstk 240
 261 PANKSENTOSLKLKLYDSPAQRYSSESDGLVFDPAKTIISRTPNNGVALPHGDHYHPTPSK 320
 241 pankse ntol slklklydspaqr yssesdg l vfdpak tiisrt p ngvalphgdhyhptpsk 300
 321 LSALEEKIARNAVPSISGSGTSTNAKPREVYSSIGLSLSSNPSITSEKLSASDGIYFN 380
 301 lsaleeki arnavpsisgsgtstn akprevyss igls slssnpsitseklsas dgiyfn 360
 381 PKDIYEETATAYIYHGHGFHYIPKSNIGOPTIPNNSLATPPSPSLPNNPTSEKHEED 440
 361 pkdiyeet atayiyhghgfhy ipk snigopt ipnnslatppspslpnnptse kheed 420
 441 GYGFDAANRIIAEDSGFYMSGDNHNYFFKDLTEQIKAAQKHLLEYVTSKNGIDSLSS 500

|||||
 421 gygfdaanriiaedsgf ymsgdnhnyffkdlteeqikaaqkhlleyvtskngidslss 480
 501 HEOPYGNAKMKKLDKTEEKIAGIMKQGVKRESTVYNNKXNATYPRGDHHAADID 560
 481 heopygnakmkldk teekia gimkqgvkre stvynn kxnatyp rgdhhaadid 540
 561 EHKPVGISHSHNTELFEPREGVAKKEGNKYTGEBELTNVNLKNSFNNQNTFLANGQ 620
 541 ehkpv gishshntelfe p regvakkegnkytg ebeltnvnlknsfnnqntflangq 600
 621 KRVSFEPPELEKKIGIMLVKLTTPDGKYLEKVSAGYGVGGINANFELDQYLPQGT 680
 601 krvsfeppelekkigimlvk ltpdgkylekvs agygv gginanf eldqylpqgt 660
 681 FKYTASADYPEVSGDGFPTSLAYKASOTIFYPHADDTLRVAPAVKGTDAL 740
 661 fkytas ady p esgdgfptsl aykasot ifyp haddtlr v apavkgt dal 720
 741 VRVDFEFGNAYLENNKYVGEIKLPIRLNQTTRTAGNKIPYFMANAYLDNSTYIVE 800
 721 vr vdfefgnayle nnyk vgeikl pirlnqttr tagnk i pyfmanayldnstyive 780
 801 VPILEKENQTRKPSILPOFKRNKAQENSKLDEKVEBPTSKVEKEKISFGNSTSNLT 860
 781 vpilekenqtrkpsilpofk rnk aqen sklde kvebptskveke k isfgnstsnlt 840
 861 EEVPTVDYVQKVAKFASVGMKLENYLFENDGTIELYPSGEYIKKMAADFTGBAPGN 920
 841 eevptvd yvqkva kfasv gmkle ny lf endgtiel ypsgey i kkm aadftg bapgn 900
 921 GENRSENGKYSTGVENQPTENRPADELPAPEKPYKPNSTDNGLNDEGNVSDPM 980
 901 genrse ngkystgv enqpt enrpa delpa pe kpykpn stdn glnde gnv sdp m 960
 981 LDPALKEAPANDYQOEKIEKFTASYGGLSDVITRMDGTIELRPSGVIYKKNISDPIA 1039
 961 ldpaleeap andy qoe k iek fta syg glsdv itrmdgtiel rpsgvi y kkn isd pia 1019

RESULT 7

AA012752
 ID AA012752 standard; Protein; 1019 AA.

AA012752;

21-NOV-2000 (first entry)

Streptococcus pneumoniae strain P421 BVH-3 protein antigen.

Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 prophylaxis; therapy; infection; diagnosis; meningitis; bacteremia;
 otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.

MO200039299-A2.

06-JUL-2000.

20-DEC-1999; 99MO-CA01218.

23-DEC-1998; 98US-0113800.

(BIOC-) BIOCHEM PHARMA INC.

Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

WPI; 2000-452397/39.

Streptococcal antigens useful for vaccinating against e.g. meningitis,
 otitis media, bacteremia and/or pneumonia.

15 Disclosure: Fig 11; 106bp; English.

16 The present invention describes nucleic acids (I) encoding protein
17 antigens (II) from Streptococcus pneumoniae. The protein antigens
18 have bactericidal activity. The nucleic acids, encoding the protein
19 antigens, may be used for the recombinant production of the proteins
20 they encode. The protein antigens may then be used as vaccines for the
21 prevention and treatment of Streptococcal infections in mammals
22 (especially humans) which result in, e.g. meningitis, otitis media,
23 bacteraemia and/or pneumonia. The present sequence represents a
24 S. pneumoniae BVH-3 protein antigen, from the present invention.

25 Sequence 1019 AA;

Query Match 97.9%; Score 5291; DB 21; Length 1019;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1015; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 21 CAYALNHRSGENKNNRVSYSOSOSKSENTPPOVOSKEIQAOIYIKITDQGY 80
2 1 cayalnhrsgenknnrvsysdsqsgskentlppdqsgkqiaeqivikltdqgyv 60
3 TSHGHHYHYNGKVPYDALFSEELMKDPNTOLKADIVNEKGGYIIRVDGKYVYLKD 140
4 61 tshghhyhyngkvpdydalffseellmkdpnylkdadlvnekkgylikvdgkyvylkd 120
5 141 AAHANVKTKEINQOKDHYKDNKVNANVAASOGRTTNDGYVNPADITDGTNA 200
6 121 aahadvktkeidnqkqghvdkvsnvavarqgyrlyndgyvnpaditdgtna 180
7 201 YVPHGHHYHYKPSDLSASELAAKAAHLAGKNNOPOLSSTSTSDNNQTSVAKGTSK 260
8 181 ylvphghhyhykpsdlsaseelaakahlagknnopsqslstsdannqtsvakgtsk 240
9 261 PANKSENTQSLKELYDPSAORSESGLVFPDPAKITSRTPNGALPHGHHYHYPSK 320
10 241 pankeentqslkelydpsaorysesglvfpdpakitsrtpngalphghhyhytprsk 300
11 321 LSALEEKARWPIPGTSTYSTNAKPNHYVSSLSGSSNSSTSTSELSASAGTYFN 380
12 301 lsaleekarwpipgtstystnkpnhyvsslsagssnsststselesasagtyfn 360
13 381 PRDIYEETATAYIYHSGHGFHYIPRSNOIGOPTLPNNSLAPSPSLPFPNGSHEKHEED 440
14 361 pxdiveetatyiyhsgghfhyiprsnigoptlpnnslapspslpfpngshekheed 420
15 441 GYGDPANRIIIEDESGFVMSHGDHNYFFKDLTEEOIKAAOKHLEEKTSHNGDLSLS 500
16 421 gygdpanriiiedesgfvmsghdnhnyffkdlteeqkkaeqkheektsahngdlsas 480
17 501 HEODYPGAKEMKDLKIEKIKAGIMQYGVKRESIYVNEKNAIYIPHGDHHDHPID 560
18 481 heodypgakemkdlkiekikagimqygvkresiyvneknaiyipghdhdhpid 540
19 561 EHKPRGIGHSHSNELFPEBEGVAKKEGNKYTGEBELTNVNLKNSFNNOONTLANQ 620
20 541 ehkprgighshsnelfpeebegvakkegnkytgebeltnvnllknsfnnoontlanq 600
21 621 KRVSFSPPELEKIKGIMLAKLITPDGKVEKYSKGKFGVGNIAFELDOPILPQOT 680
22 601 krvsfsppelekiikgimlavltpdgkvekyksgkfgvgvgniafelldopilpqot 660
23 681 FKYTASADYEVSTDGTFYPTSLAYMMASQITTFPHADITLKNVPOPAVPKGTAL 740
24 661 fkytasadyevstdgtyfptslaymmasqitfphaditlknvpoavpkgtal 720
25 741 VRVDEFGNAYLIENTVVGKIKLIPKLNGTRTAGNKIPVEMANAYADNOSTYVE 800
26 721 vrvdefgnaylienrvvgkiklipklngtrtagnkipvemaayadnostyve 780
27 801 VPILEKENQTKPSLIPQFKRNKAQENSKLDEKVEEPTSEKVEKLSGNGSTNSTL 860

DB 781 vpilekenqtkpslppqfkrnkagenskidekveeptsekekssetgnstnstl 840
QY 861 EEVPPVDPVOEKVAKFASVAMKLENTLFNMDGTELYLPGEVYIKKMMAPFTGAPQGN 920
DB 841 eeppvdpvokvafasvaymklenylnmdgtelylpsevalkmmadrtgeapqgn 900
QY 921 GENPSENGKYSTGVENQPTENRPAUSLPEAPNEKPYKPNSTDNGLNDEGNVGSDDPM 980
DB 901 genpsengkystgvtenqptenrpadlpeapnekpykpnstndnglnpegnvgsdpm 960
QY 981 LDPALEAPVADPVOEKLEKFTASTAGLDSVIFNMOSTITLRLPGEVYIKKNSDPTA 1039
DB 961 ldpaaleapvadvpveklekftastagldsvifnmogtltlrlpsgevalkknisdlla 1019

RESULT 8

AB12753
ID AB12753 standard; Protein; 1019 AA.

AC AB12753;

DF 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae; strain A66 BVH-3 protein antigen.

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

KW otitis media; pneumonia; immunisation; bactericidal.

OS Streptococcus pneumoniae.

PN WO200039299-A2.

XX 06-JUL-2000.

XX 20-DEC-1999; 99WO-CA01218.

XX 23-DEC-1998; 98US-0113800.

XX (BIOC-) BIOCHEM PHARMA INC.

PI Hamel J, Brodeur BR, Plneau I, Martin D, Rioux C, Charland N;

XX WPI; 2000-452397/39.

XX Streptococcal antigens useful for vaccinating against e.g. meningitis,

PT otitis media, bacteraemia and/or pneumonia -

PS Disclosure: Fig 11; 106bp; English.

CC The present invention describes nucleic acids (I) encoding protein

CC antigens (II) from Streptococcus pneumoniae. The protein antigens

CC have bactericidal activity. The nucleic acids, encoding the proteins

CC antigens, may be used for the recombinant production of the proteins

CC they encode. The protein antigens may then be used as vaccines for the

CC prevention and treatment of Streptococcal infections in mammals

CC (especially humans) which result in, e.g. meningitis, otitis media,

CC bacteraemia and/or pneumonia. The present sequence represents a

CC S. pneumoniae BVH-3 protein antigen, from the present invention.

XX Sequence 1019 AA;

Query Match 97.9%; Score 5291; DB 21; Length 1019;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1015; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 21 CAYALNHRSGENKNNRVSYSOSOSKSENTPPOVOSKEIQAOIYIKITDQGY 80
DB 1 cayalnhrsgenknnrvsysdsqsgskentlppdqsgkqiaeqivikltdqgyv 60
QY 81 TSHGHHYHYNGKVPYDALFSEELMKDPNTOLKADIVNEKGGYIIRVDGKYVYLKD 140

61 tshgdyhyngkvydalfseellmkdpnyqkdaadivevkggylikvdgkyxyylkd 120
Y AAADNVRTDEINROKOEHVKNDEKYNVAVASQGRYTTNDGYFNPADIIETDGN 200
b 121 aahadavrttdelnrqgvehkdnkynsnvavargryttndgyfnpadlledgna 180
Y YIVPHGHHYIIPKSDISASELAAKAHLAKKNQPSQLSYSTASDNNTQSAKGSTK 260
b 181 yivphghyhyipkdsiaselaaakahlagknmpsqqlsyssasdnntqsvakgstsk 240
Y PANKSEMSLSEKLELSPASQORYSESDGLVDPAKIISTPNCVAILPHGDHYHFIYSK 320
b 241 panlsenlgellkelyspasqoryseesdglyvdpaklilstpncvailphgdhyhfiysk 300
Y LSALAEKRIARVPISGTSVSTNAKPNEVVSLSGLSSNPSSLTTSKELSSASDGYTFN 380
b 301 lsaleekriarvpisgtsvstnakpnevvsalsglssnpsalttskelssasdgylfn 360
Y PKDIVEETATAYIVRHGDHRYIIPKSNQIGOPTLPNNSLATPSPSLPINTGSHKEHEED 440
b 361 pkdiveetateayivrhgdhryipksnqigoptlpnnslatpsslpintgstshkheed 420
Y GYGDFANRIIAEDSGVYMSGDHNFYFKKDLTEBOIKAAOKLEKVTSHNGLSLSS 500
b 421 gygdfanriiaedsgvymsgdhnfyfkldteeqkkaqkleeekshnglsdiss 480
Y HEODIPGNAKEMKDLKRIEIKINGIKOYGVKNESIVVKNKNAIITYPHGDHHAAPID 560
b 481 heodipgnakemkdlkrieikngikoygvknesivvknknaiityphgdhhaapid 540
Y EHKRVGIGHSHSNELKRPBGVAKKRGKNVYTGSELTNNVNLKSTPNQNTLANCO 620
b 541 ehkrvgighshsnelkrpbgvakrgknvytgseeltnnvnlkstpnqntlanng 600
Y KRVSFSPPELEKELGIMLYKLTTPDGKYLEKSGVFGVGNINLEFDQCYLGGOT 680
b 601 krvsfsppelekelligimlyklttpdgkyleksgvfgvgninlefdqcyllggot 660
Y EKYTIASKDPREVSIDGTFFVPSLAYKMASQITFFPHAGDYTLRVNPPFAVKGTDAL 740
b 661 ekytiaskdprevsidgtffvpslaykmasqitffphagdytlrvnppfavykgtdal 720
Y VRVDEHNGAAYLENNKTVGEIKPIPKLNGOTTRTGNKIPVTFMANAYLDNOSTIIVE 800
b 721 vrvdehngaaaylennktvgeikpipklnngottrtgnkipvtfmanayldngostive 780
Y VPILEKNOGDKPSILFOFRKNKQENSKLDEKVEEPTSKVEKELSESTNSSTL 860
b 781 vpileknogdkpsilfofrknkqenskldkveeptskvekelsestnsstl 840
Y EEPVTPVQEKVAKFAESYGMKLENTLFNMDGTIELYLPSEGVYIKKNMADFTGEAOGN 920
b 841 eepvtpvqekvakfaesygmkleentlfnmddgtielylpsegvikknmadftgeaogn 900
Y GENRPSNGKYSTGTVENOTFENKPADSLPEAPNEKRYKENSNDKNLMEGAVGSDPM 980
b 901 genrpsngkystgtvenotfennkpadslpeapnekrykenskndknlmepegavgsdpm 960
Y LDPALAEAPAVDPQOEKLEKFTASYGGLDSVIFNMMDGTIELRLPSEGVYIKKNLSDTIA 1039
b 961 ldpalaeapavdpqoelekelftasysgglsvifnmddgtielylrpsegvikknlsdtia 1019

ESTUT 9
D AAB12721 standard; Protein: 840 AA
X AAB12721:
X 21-NOV-2000 (first entry)
X Streptococcus pneumoniae SP63 BVH-3 protein antigen SEQ ID NO:16.
X

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW Propylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW Otitis media; pneumonia; immunisation; bactericidal.
OS Streptococcus pneumoniae.
PN MO200039299-A2.
PD 06-JUL-2000.
PF 20-DEC-1999; 99MO-CA01218.
PR 23-DEC-1998; 98US-0113800.
XX (BIOC-) BIOCHEM PHARMA INC.
PI Hamel J, Brodeur BR, Pneau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
DR N-PSDB; AAA65738.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PS Otitis media, bacteraemia and/or pneumonia -
XX Claim 16; Fig 19; 106pp; English.
XX
CC The present invention describes nucleic acids (i) encoding protein
CC antigens (ii) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae SP63 BVH-3 protein antigen.
XX
SQ Sequence 840 AA:

Query Match 79.1%; Score 4276.5; DB 21; Length 840;
Best Local Similarity 82.3%; Pred. No. 4e-251;
Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;

QY 21 CAYALNORSEKNNKRVSYVDGSSOSKSENITPVOVQKSGIAEOIVYKITDGYT 80
b 1 cayalngqrsenkdnnrvsyvdgssosksenitpovqksqiaeqivikltdgyv 60
QY 81 TSHGDHYHYNGVVPYDALFSEELMKDPNYQKDAADIVEVKGYYIIRYDGKYYVYLKD 140
b 61 tshgdyhyngkvydalfseellmkdpnyqkdaadivevkggylikvdgkyxyylkd 120
QY 141 AAADNVRTDEINROKOEHVKNDEKYNVAVASQGRYTTNDGYFNPADIIETDGN 200
b 121 aahadavrttdelnrqgvehkdnkynsnvavargryttndgyfnpadlledgna 180
QY 201 YIVPHGHHYIIPKSDISASELAAKAHLAKKNQPSQLSYSTASDNNTQSAKGSTK 260
b 181 yivphghyhyipkdsiaselaaakahlagknmpsqqlsyssasdnntqsvakgstsk 240
QY 261 PANKSEMSLSEKLELSPASQORYSESDGLVDPAKIISTPNCVAILPHGDHYHFIYSK 320
b 224 ----- 223
QY 321 LSALAEKRIARVPISGTSVSTNAKPNEVVSLSGLSSNPSSLTTSKELSSASDGYTFN 380
b 224 ----- 223
QY 381 PKDIVEETATAYIVRHGDHRYIIPKSNQIGOPTLPNNSLATPSPSLPINTGSHKEHEED 440
b 224 ----- 243
QY 441 GYGDFANRIIAEDSGVYMSGDHNFYFKKDLTEBOIKAAOKLEKVTSHNGLSLSS 500
b -----

D AAB12731 standard; Protein: 568 AA.
 X AAB12731;
 X
 X 21-NOV-2000 (first entry)
 X Streptococcus pneumoniae NEW1 protein antigen SEQ ID NO:64.
 X Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 X prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 X otitis media; pneumonia; immunisation; bactericidal.
 X Streptococcus pneumoniae.
 X MO200039299-A2.
 X
 X 06-JUL-2000.
 X
 X 20-DEC-1999; 99MO-CA01218.
 X
 X 23-DEC-1998; 98US-0113800.
 X (BIOC-) BIOCHEM PHARMA INC.
 X Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 X WPI; 2000-452397/39.
 X Streptococcal antigens useful for vaccinating against e.g. meningitis,
 X otitis media, bacteraemia and/or pneumonia -
 X
 X Claim 18; Fig 29; 106pp; English.
 X
 X The present invention describes nucleic acids (I) encoding protein
 X antigens (II) from Streptococcus pneumoniae. The protein antigens
 X have bactericidal activity. The nucleic acids, encoding the proteins
 X antigens, may be used for the recombinant production of the proteins
 X they encode. The protein antigens may then be used as vaccines for the
 X prevention and treatment of Streptococcal infections in mammals
 X (especially humans) which result in, e.g. meningitis, otitis media,
 X bacteraemia and/or pneumonia. The present sequence represents the
 X S. pneumoniae NEW1 protein antigen.
 X
 X Sequence 568 AA;
 Q

Query Match 54.6%; Score 2953; DB 21; Length 568;
 Best Local Similarity 100.0%; Pred. No. 3.8e-171;
 Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 472 DLTEEOIKAAQKHEEYKTSNGDSLSSEHODYPGNAKEMKLDKRIEKGIMKQY 531
 |||||||
 b 1 dlteeqkaeqkhlheektsngdsissheqdygnakemkldkriekingimkqys 60
 |||||||
 Y 532 VKRESIVYNNKNAIIPDGDHDAIDDEKPVGICHSSNVELFPEEGVAKKEKNV 591
 |||||||
 b 61 vkresivnnknaaiipdgdhdaidekpvigichssnvelfpeegvakkeknv 120
 |||||||
 Y 592 YTGELNYYNVLKNSFNNOFLANGOKRVSPFPELEKKGIMLVKLTIPDGRVL 651
 |||||||
 b 121 ytgelnnyvnlknsfnnoflangokrvspfpelkkkgimlvkltipdgrvl 180
 |||||||
 Y 652 EKVSQVFGSGVGNIANFELDQPTLPQTFKTYTASNDYEVSTGTFYPTSLAYKMAS 711
 |||||||
 b 181 ekvsqvfsgvgvgnianfelldqptlpqtfktytasndyevstgtyfptslaykmas 240
 |||||||
 Y 712 QTIYPPHAGDTYLRNVPORAVRKGTDALYRVDEHGNALYLENNYVGEIKLPIPLNQ 771
 |||||||
 b 241 qtiypphagdtylrnvporavrkgtDALYRVDEFHGNALYLENNYVGEIKLPIPLNQ 300
 |||||||
 Y 772 GTTFATANKIIPVFNANAYDNDOSTYIVVPILEKENOTKPSILPQFKNKAOENSKLD 831
 |||||||
 b 301 gtctaaqkhlheektsngdsissheqdygnakemkldkriekingimkqys 360
 |||||||

QY 832 EKVEEPTSEKVEKEKLSSTNGSTNSLTLEEVPTDVOEKVAFASYSKMLENVLPM 891
 |||||||
 Db 361 ekveeptsekevkeklsstngstnsltleevptdvoekvafasykmlelvpm 420
 |||||||
 QY 892 DGTIELYPSGEVTKKMAADFTGAPGNGENKPSNGKYSTGVENQPIENKPADSLPE 951
 |||||||
 Db 421 dgtielypsgevlkkmadftgeapngngenkpsngkystgvnqpienkpadslpe 480
 |||||||
 QY 952 APNEKPYKPNSTNDGKLNDEGNVGSDDPMDPALEAPAVDPQEKLEKFTASVGLSDS 1011
 |||||||
 Db 481 apnekpykpnstndgklndegnvgsddpmdpaleapavdpqeklekftasyglids 540
 |||||||
 QY 1012 VIFNMDGTIELRLPSGEVTKKNSDFFIA 1039
 |||||||
 Db 541 vifmngdtielyrlpsgevlkknlsdffa 568
 |||||||

RESULT 12

AAB12725
 ID AAB12725 standard; Protein: 1057 AA.
 X
 X AAB12725;
 X
 X 21-NOV-2000 (first entry)
 X Streptococcus pneumoniae NEW12 protein antigen SEQ ID NO:58.
 X
 X Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 X prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 X otitis media; pneumonia; immunisation; bactericidal.
 X
 X Streptococcus pneumoniae.
 X
 X MO200039299-A2.
 X
 X 06-JUL-2000.
 X
 X 20-DEC-1999; 99MO-CA01218.
 X
 X 23-DEC-1998; 98US-0113800.
 X (BIOC-) BIOCHEM PHARMA INC.
 X Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 X WPI; 2000-452397/39.
 X Streptococcal antigens useful for vaccinating against e.g. meningitis,
 X otitis media, bacteraemia and/or pneumonia -
 X
 X Claim 18; Fig 23; 106pp; English.
 X
 X The present invention describes nucleic acids (I) encoding protein
 X antigens (II) from Streptococcus pneumoniae. The protein antigens
 X have bactericidal activity. The nucleic acids, encoding the proteins
 X antigens, may be used for the recombinant production of the proteins
 X they encode. The protein antigens may then be used as vaccines for the
 X prevention and treatment of Streptococcal infections in mammals
 X (especially humans) which result in, e.g. meningitis, otitis media,
 X bacteraemia and/or pneumonia. The present sequence represents the
 X S. pneumoniae NEW12 protein antigen.
 X
 X Sequence 1057 AA;
 SQ

Query Match 54.6%; Score 2953; DB 21; Length 1057;
 Best Local Similarity 100.0%; Pred. No. 9e-171;
 Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEOIKAAQKHEEYKTSNGDSLSSEHODYPGNAKEMKLDKRIEKGIMKQY 531
 |||||||
 Db 1 dlteeqkaeqkhlheektsngdsissheqdygnakemkldkriekingimkqys 60
 |||||||

Y 532 VKRSIIYVKKENNAIIPHGHHADPIDEHKPVGIGHSHSNYTELFKEPSEVAKKEGKV 591
 |||||
 Y 61 VKRESIYVKKENNAIIPHGHHADPIDEHKPVGIGHSHSNYTELFKEPSEVAKKEGKV 120
 |||||
 Y 592 YTGSELINVVNLKNSFTFNNONFTLANQKRVSSFPPELKKIGINLVLTIPDGKVL 651
 |||||
 Y 121 YTGSELINVVNLKNSFTFNNONFTLANQKRVSSFPPELKKIGINLVLTIPDGKVL 180
 |||||
 Y 652 EKVSGRKVGEEVGINANFELQPIIPGOTFEKTTASKDPEVYSIDGFTVTSTAYKAS 711
 |||||
 Y 181 EKVSGRKVGEEVGINANFELQPIIPGOTFEKTTASKDPEVYSIDGFTVTSTAYKAS 240
 |||||
 Y 712 OTIYFPFAGGTLYRVNQFVAPKGTDLVRFDEFGHNALENNYKVGELKLPKILNO 771
 |||||
 Y 241 QTLIYPIHAGTLYRVNQFVAPKGTDLVRFDEFGHNALENNYKVGELKLPKILNO 300
 |||||
 Y 772 GTRTAGNKRIFVTMANAYLDNOSTYIYVEPILKEKNOTDPSILPQKRRKADENSKLD 831
 |||||
 Y 301 GTTRAGNKRIFVTMANAYLDNOSTYIYVEPILKEKNOTDPSILPQKRRKADENSKLD 360
 |||||
 Y 832 EKVEEPTSEVVEKEKSESTNSNSTLEEVPTVDPVQEKVAKFAESTYGMKLENVLFNM 891
 |||||
 Y 361 EKVEEPTSEVVEKEKSESTNSNSTLEEVPTVDPVQEKVAKFAESTYGMKLENVLFNM 420
 |||||
 Y 892 DGTIELYLPSEGVIRKKNADFTGEAPOGNGENKPSENGKVGSTGVENQPTENKPADSLPE 951
 |||||
 Y 421 DGTIELYLPSEGVIRKKNADFTGEAPOGNGENKPSENGKVGSTGVENQPTENKPADSLPE 480
 |||||
 Y 952 APNKKPVKPNSTNGMLNPGNVGSDPMLDPALEAPVADPVQEKLEKFTASYGLGIDS 1011
 |||||
 Y 481 APNKKPVKPNSTNGMLNPGNVGSDPMLDPALEAPVADPVQEKLEKFTASYGLGIDS 540
 |||||
 Y 1012 VIFNMDSGTEIRLSESGEVIKKNLSDFTA 1039
 |||||
 Y 541 VIFNMDSGTEIRLSESGEVIKKNLSDFTA 568
 |||||
 RESULT 13
 ID AAB12719 standard; Protein: 528 AA.
 AAB12719;
 21-NOV-2000 (first entry)
 Streptococcus pneumoniae BVH-3B protein antigen SEQ ID NO:10.
 Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 M prophylaxis; therapy; infection; diagnosis; meningitis; bacteremia;
 W otitis media; pneumonia; immunisation; bactericidal.
 Streptococcus pneumoniae.
 NC0200039299-A2.
 06-JUL-2000.
 20-DEC-1999; 99NC-CA01218.
 23-DEC-1998; 98US-0113800.
 (BIOC-) BIOCHEM PHARMA INC.
 Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 MPI; 2000-452397/39.
 N-PSDB; AAA65734.
 Streptococcal antigens useful for vaccinating against e.g. meningitis,
 otitis media, bacteremia and/or pneumonia -
 Claim 18; Fig 10; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteremia and/or pneumonia. The present sequence represents the
 CC S. pneumoniae BVH-3B protein antigen.
 XX
 SO Sequence 528 AA;
 Query Match 50.8%; Score 2746; DB 21; Length 528;
 Best Local Similarity 100.0%; Pred. No. 1,2e-158; Indels 0; Gaps 0;
 Matches 528; Conservative 0; Mismatches 0;
 Y 512 MKDLDRKTEERKTIAGIMKQYGVKRSIYVKKENNAIIPHGHHADPIDEHKPVGIGHSH 571
 |||||
 Y 1 MKDLDRKTEERKTIAGIMKQYGVKRSIYVKKENNAIIPHGHHADPIDEHKPVGIGHSH 60
 |||||
 Y 572 SNTELEKPEEVEVAKKEGKVYTGSELINVVNLKNSFTFNNONFTLANQKRVSSFPPEL 631
 |||||
 Y 61 SNTELEKPEEVEVAKKEGKVYTGSELINVVNLKNSFTFNNONFTLANQKRVSSFPPEL 120
 |||||
 Y 632 EKKIGINLVLTIPDGKVLKEKVGKVGEEVGINANFELQPIIPGOTFEKTTASKDYP 691
 |||||
 Y 121 EKKIGINLVLTIPDGKVLKEKVGKVGEEVGINANFELQPIIPGOTFEKTTASKDYP 180
 |||||
 Y 692 EVSYDGTFTVPTSLAYKKAQOTIYFPFHAGDTLYRVNQFVAPKGTDLVRFDEFGHNA 751
 |||||
 Y 181 EVSYDGTFTVPTSLAYKKAQOTIYFPFHAGDTLYRVNQFVAPKGTDLVRFDEFGHNA 240
 |||||
 Y 752 YLENNYKVGELKLPKILNOGTRTAGNKRIFVTMANAYLDNOSTYIYVEPILKEKNOTD 811
 |||||
 Y 241 YLENNYKVGELKLPKILNOGTRTAGNKRIFVTMANAYLDNOSTYIYVEPILKEKNOTD 300
 |||||
 Y 812 KPSILPQKRRKADENSKLDKEVEEPTSEVVEKEKSESTNSNSTLEEVPTVDPVQEK 871
 |||||
 Y 301 KPSILPQKRRKADENSKLDKEVEEPTSEVVEKEKSESTNSNSTLEEVPTVDPVQEK 360
 |||||
 Y 872 KVAKFAESTYGMKLENVLFNMDSGTEIRLSESGEVIKKNADFTGEAPOGNGENKPSENGKV 931
 |||||
 Y 361 KVAKFAESTYGMKLENVLFNMDSGTEIRLSESGEVIKKNADFTGEAPOGNGENKPSENGKV 420
 |||||
 Y 932 STGTVENQPTENKPADSLPEAPNKKPVKPNSTNGMLNPGNVGSDPMLDPALEAPV 991
 |||||
 Y 421 STGTVENQPTENKPADSLPEAPNKKPVKPNSTNGMLNPGNVGSDPMLDPALEAPV 480
 |||||
 Y 992 DPVQEKLEKFTASYGLGIDSYTFNMDSGTEIRLSESGEVIKKNLSDFTA 1039
 |||||
 Y 481 DPVQEKLEKFTASYGLGIDSYTFNMDSGTEIRLSESGEVIKKNLSDFTA 528
 |||||
 RESULT 14
 ID AAB12724 standard; Protein: 509 AA.
 AAB12724;
 21-NOV-2000 (first entry)
 Streptococcus pneumoniae L-BVH-3-AD protein antigen SEQ ID NO:57.
 Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteremia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 Streptococcus pneumoniae.
 NC0200039299-A2.

D 06-JUL-2000.
 X 20-DEC-1999; 99MO-CA01218.
 F X
 X 23-DEC-1998; 98US-0113800.
 X
 X (BIOC-) BIOCHEM PHARMA INC.
 X Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N,
 X WPI; 2000-452397/39.
 X Streptococcal antigens useful for vaccinating against e.g. meningitis,
 X otitis media, bacteremia and/or pneumonia.
 X
 X Claim 18; Fig 22; 106pp; English.
 X
 X The present invention describes nucleic acids (I) encoding protein
 C antigens (II) from Streptococcus pneumoniae. The protein antigens
 C have bactericidal activity. The nucleic acids, encoding the protein
 C antigens, may be used for the recombinant production of the proteins
 C they encode. The protein antigens may then be used as vaccines for the
 C prevention and treatment of Streptococcal infections in mammals
 C (especially humans) which result in, e.g. meningitis, otitis media,
 C bacteremia and/or pneumonia. The present sequence represents the
 C S. pneumoniae L-BVH-3-AD protein antigen.

Sequence 509 AA;

Query Match 49.0%; Score 2650; DB 21; Length 509;
 Best Local Similarity 100.0%; Pred. No. 7, 6e-153;
 Matches 509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKRSKYIAGSAVYISLCAYALNHSOEKNNRVSYVDSOSQSENIPTPOVS 60
 1 mkrskyiaagsavylsalcayalnhsrgeknrvsyvdsqsgsqsensltpdqv 60
 Y OKSGIOAEQIVITITDOGYTSHGDHYHNGVYPYDALPSEELMKDPYQKADIVN 120
 b 61 qksgioaeqivlitdogytshgdhyhngvypydalseelmkdpdyqkadvln 120
 Y 121 EVKGGYIIKVDGKYYTYLKAADNRVTKDEINRKQEHYKNEKYNVAVARSGRY 180
 b 121 evkgyyllkvdgkytyyllkaadnrvtkdelnrqgehykneynsvavarsgry 180
 Y 181 TTNDGVVFNPAIDIEPTGNAYITPBGHGYHYIKPSDLSASELAAKAHLAGKMPQSOLS 240
 b 181 ttndgvvfnpadiedetgnayitpbgghyhyipksdlsaseelsaahlagkmpqsols 240
 Y 241 YSSTADNNTQSAVAKSTSPANKSENLSILKELDPSAQRYSSESDGLVFPKATISR 300
 b 241 ysstadnntqsaavakstspanksenlsilkelypsraqryssedglvfpaktisr 300
 Y 301 TPNGVAIPBGDHYHTIPYKSLALEEKIARMPISGSGTVSNAPNEVYSSLSGLSSN 360
 b 301 tpngvailpbgdhyhtipykslaleekiarmplisgsgtvsnaapnevyslsglssn 360
 Y 361 PSSLTQSKELSSADGYIFNPKDIVEETATAYIVRHGDHHTPKNOIGOPTLPNNSLA 420
 b 361 pssltqskelssadgylfnpkdiveetacayivrngdhytpkbnqlyqpcilpnnsia 420
 Y 421 TPSPSLPINTGTSHEHHEDEGYGFDMNRITAEDESGFVMSHGDHNYFFKKDTEEQIKR 480
 b 421 tpspslpintgtshehheedegygfdmnrityaedesgfvmshgdhnyffkkdteeqika 480
 Y 481 AQKHEAEVTSKSHGLDLSLSEHEDYPCNA 509
 b 481 aqkheaevtshngldslseshedgypna 509

RESULT 15
 AB12723

ID AAB12723 standard; Protein; 489 AA.
 XX
 AC AAB12723;
 XX
 DT 21-NOV-2000 (first entry)
 DE
 DE Streptococcus pneumoniae BVH-3AD protein antigen SEQ ID NO:56.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteremia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN MO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 XX 20-DEC-1999; 99MO-CA01218.
 XX
 XX 23-DEC-1998; 98US-0113800.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N,
 XX
 PI WPI; 2000-452397/39.
 XX
 DR Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteremia and/or pneumonia.
 PT
 PS Claim 18; Fig 21; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteremia and/or pneumonia. The present sequence represents the
 CC S. pneumoniae BVH-3AD protein antigen.
 CC
 SQ Sequence 489 AA;
 Query Match 47.3%; Score 2559; DB 21; Length 489;
 Best Local Similarity 100.0%; Pred. No. 2, 4e-147;
 Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

21 CAYALNHSOEKNNRVSYVDSOSQSENIPTPOVSOKSGIOAEQIVITITDOGYV 80
 1 cayalnhsrgeknrvsyvdsqsgsqsensltpdqvqksgiaeqivlitdogyv 60
 Y 81 TSHGDHYHNGVYPYDALPSEELMKDPYQKADIVNEVKGIIYKVDGKYYTYLKD 140
 b 81 tshgdhyhngvypydalseelmkdpdyqkadvnevkgyyllkvdgkytyyllkd 140
 Y 141 AAADNRVTKDEINRKQEHYKNEKYNVAVARSGRTTNDGVVFNPAIDIEPTGN 200
 b 141 aaadnrvtkdelnrqgehykneynsvavarsgryttndgvvfnpadiedetgna 180
 Y 201 YIYPBGHGYHYIKPSDLSASELAAKAHLAGKMPQSOLSYSTADNNTQSAVAKSTSK 260
 b 201 yiypbgghyhyipksdlsaseelsaahlagkmpqsolyssstadnntqsaavakstsk 240
 Y 261 PANKSENLSILKELDPSAQRYSSESDGLVFPKATISRTPNGVAIPBGDHYHTIPYK 320
 b 261 panksenlsilkelypsraqryssedglvfpaktisrtpngvailpbgdhyhtipyk 300
 Y 321 LSALEEKIARMPISGSGTVSNAPNEVYSSLSGLSSNPSLTQSKELSSADGYIFN 380
 b 301 lsaleekiarmplisgsgtvsnaapnevyslsglssnpsltqskelssadgylfn 360

QY 381 PKDVEETATAYIVRHGDHPHYIPKSNQIGOPTLPNNSLATPSPSLP INPQTSHEKHEED 440
Db 361 pkdlveetaataylvrhgdhhyipksnqigoptlpnnslatpsslp inpqtshkheed 420
QY 441 GYGF DANRIIAEDSGFVMSGHGDNHYFFKKDLTEQIKAAQKHLEEVKTSNGLDLSLS 500
Db 421 gygf danriiaedesgfvmshgdnhyffkkdlteeqikaqkhleevktsnngldels 480
QY 501 HEQDYPGNA 509
Db 481 heqdypgna 489

Arch completed: September 26, 2001, 22:30:29
Job time: 3208 sec

SEP 27 2001 09:15:57

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 22:30:29 : Search time 57.46 Seconds

(without alignments)
143,489 Million cell updates/sec

Title: US-09-471-255-2_COPY_15_150

Perfect score: 711
Sence: 1 IVSLSLCAVYALNHRSEOK.....DGKRYVYLNKAHADNVRTRK 136

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /SIDSL/gcgdata/geneseq/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/AA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description |
|------------|-------|-------------|-----------|----|--------------------|
| 1 | 711 | 100.0 | 484 | 21 | Streptococcus pneu |
| 2 | 711 | 100.0 | 484 | 21 | Recombinant Varian |
| 3 | 711 | 100.0 | 484 | 21 | Streptococcus pneu |
| 4 | 711 | 100.0 | 485 | 21 | Streptococcus pneu |
| 5 | 711 | 100.0 | 509 | 21 | Streptococcus pneu |
| 6 | 711 | 100.0 | 1039 | 21 | Streptococcus pneu |
| 7 | 687 | 96.6 | 205 | 21 | Streptococcus pneu |
| 8 | 687 | 96.6 | 489 | 21 | Streptococcus pneu |
| 9 | 687 | 96.6 | 780 | 21 | Streptococcus pneu |
| 10 | 687 | 96.6 | 840 | 21 | Streptococcus pneu |
| 11 | 687 | 96.6 | 1019 | 21 | Streptococcus pneu |

| | | | | | |
|----|-------|------|------|----|--------------------|
| 12 | 687 | 96.6 | 1019 | 21 | Streptococcus pneu |
| 13 | 687 | 96.6 | 1019 | 21 | Streptococcus pneu |
| 14 | 687 | 96.6 | 1019 | 21 | Streptococcus pneu |
| 15 | 687 | 96.6 | 1019 | 21 | Streptococcus pneu |
| 16 | 687 | 96.6 | 1019 | 21 | Streptococcus pneu |
| 17 | 687 | 96.6 | 1019 | 21 | Streptococcus pneu |
| 18 | 687 | 96.6 | 1019 | 21 | Streptococcus pneu |
| 19 | 605 | 85.1 | 840 | 21 | Streptococcus pneu |
| 20 | 600.5 | 84.5 | 827 | 21 | Streptococcus pneu |
| 21 | 599.5 | 84.3 | 826 | 21 | Streptococcus pneu |
| 22 | 599.5 | 84.3 | 838 | 21 | Streptococcus pneu |
| 23 | 599.5 | 84.3 | 838 | 21 | Streptococcus pneu |
| 24 | 596 | 83.8 | 690 | 21 | Streptococcus pneu |
| 25 | 596 | 83.8 | 690 | 21 | Streptococcus pneu |
| 26 | 596 | 83.8 | 821 | 21 | Streptococcus pneu |
| 27 | 596 | 83.8 | 821 | 21 | Streptococcus pneu |
| 28 | 592 | 83.3 | 819 | 21 | Streptococcus pneu |
| 29 | 589.5 | 82.9 | 819 | 21 | Streptococcus pneu |
| 30 | 585.5 | 82.3 | 807 | 21 | Streptococcus pneu |
| 31 | 585.5 | 82.3 | 811 | 21 | Streptococcus pneu |
| 32 | 585.5 | 82.3 | 819 | 21 | Streptococcus pneu |
| 33 | 585.5 | 82.3 | 819 | 21 | Streptococcus pneu |
| 34 | 585.5 | 82.3 | 820 | 21 | Streptococcus pneu |
| 35 | 585.5 | 82.3 | 834 | 21 | Streptococcus pneu |
| 36 | 581.5 | 81.8 | 805 | 21 | Streptococcus pneu |
| 37 | 581.5 | 81.8 | 811 | 21 | Streptococcus pneu |
| 38 | 581.5 | 81.8 | 811 | 21 | Streptococcus pneu |
| 39 | 581.5 | 81.8 | 816 | 21 | Streptococcus pneu |
| 40 | 581.5 | 81.8 | 816 | 21 | Streptococcus pneu |
| 41 | 581.5 | 81.8 | 816 | 21 | Streptococcus pneu |
| 42 | 575.5 | 80.9 | 763 | 19 | Streptococcus pneu |
| 43 | 574.5 | 80.8 | 811 | 21 | Streptococcus pneu |
| 44 | 574 | 80.7 | 796 | 19 | Streptococcus pneu |
| 45 | 428 | 60.2 | 825 | 22 | Streptococcus pneu |

ALIGNMENTS

RESULT. 1

ID AAB12718 standard; Protein; 484 AA.

AC AAB12718;

DT 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae BVH-3A protein antigen SEQ ID NO:8.

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

KW otitis media; pneumonia; immunisation; bactericidal.

OS Streptococcus pneumoniae.

PN MO200039299-A2.

PD 06-JUL-2000.

PF 20-DEC-1999; 99MO-CA01218.

PR 23-DEC-1998; 98US-0113800.

PA (BIOC-) BIOCHEM PHARMA INC.

PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

DR WPI; 2000-452397/39.

PT N-PSDB; AA65733.

XX Streptococcal antigens useful for vaccinating against e.g. meningitis,

XX otitis media, bacteraemia and/or pneumonia

35 Claim 18; Fig 8; 106pp; English.

36 The present invention describes nucleic acids (I) encoding protein
37 antigens (II) from Streptococcus pneumoniae. The protein antigens
38 have bactericidal activity. The nucleic acids, encoding the proteins
39 antigens, may be used for the recombinant production of the proteins
40 they encode. The protein antigens may then be used as vaccines for the
41 prevention and treatment of Streptococcal infections in mammals
42 (especially humans) which result in, e.g. meningitis, otitis media,
43 bacteraemia and/or pneumonia. The present sequence represents the
44 S. pneumoniae BVH-3A protein antigen.

45 Sequence 484 AA:

Query Match 100.0%; Score 711; DB 21; Length 484;
Best Local Similarity 100.0%; Pred. No. 8.3e-70;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 IVSLSCAYALNQHRSQENKNNRVSYVDGSSQSSQSENLPDQVSOREGIOAEQIVIKI 60
15 IVSLSCAYALNQHRSQENKNNRVSYVDGSSQSSQSENLPDQVSOREGIOAEQIVIKI 74
61 TDGQVYTSBGHDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGYYIKVQDKY 120
75 TDGQVYTSBGHDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGYYIKVQDKY 134
121 VYVLLKDAHADNVRTK 136
135 VYVLLKDAHADNVRTK 150

RESULT 2
AB01467 standard; Protein: 484 AA.

AB01467;

20-OCT-2000 (first entry)

Recombinant variant of Sp36 (Sp36E) of S. pneumoniae.

Streptococcus pneumoniae; infection; vaccine; collid coll region;
histidine triad residue; Sp36; antibody; otitis media;
nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
meningitis; lobar pneumonia.

Streptococcus pneumoniae.

Key Location/Qualifiers
Region 63..68
/label= Histidine triad residue
Region 120..140
/label= Collid coll region
Region 185..190
/label= Histidine triad residue
Region 289..294
/label= Histidine triad residue
Region 376..381
/label= Histidine triad residue
Region 441..446
/label= Histidine triad residue
Region 750..772
/label= Collid coll region

W0200037105-A2.

29-JUN-2000.

21-DEC-1999; 99MO-US30390.

21-DEC-1998; 98US-0113048.

PA (MED-) MEDIMUNE INC.

PI Johnson LS, Koenig S, Adamou JE;

DR WPI: 2000-452129/39.

DR N-PSDB: AAA47603.

PT Vaccine useful for prophylaxis and treatment of pneumococcal infections
PT such as otitis media, nasopharyngeal and bronchial infections,
PT comprises Streptococcus pneumoniae proteins

Claim 1; Page 58-60; 70pp; English.

Although a number of proteins have been suggested as being involved
in the pathogenicity of Streptococcus pneumoniae, there still remains
a need to identify polypeptides having epitopes in common from
various strains of S. pneumoniae in order to utilise such
polypeptides in vaccines to protect against a wide variety of
S. pneumoniae. New vaccine compositions are described which comprise a
Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
acids in length that comprise at least one histidine triad residue
(HXHXH) or a collid-coll region, or an antibody directed against
these features. The vaccine is useful in protecting against infection
by Streptococcus pneumoniae. The vaccine composition comprising
antibodies to is useful for passive immunization for treating
pneumococcal infections which includes otitis media, nasopharyngeal
and bronchial infections.

Sequence 484 AA:

Query Match 100.0%; Score 711; DB 21; Length 484;
Best Local Similarity 100.0%; Pred. No. 8.3e-70;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 IVSLSCAYALNQHRSQENKNNRVSYVDGSSQSSQSENLPDQVSOREGIOAEQIVIKI 60
15 IVSLSCAYALNQHRSQENKNNRVSYVDGSSQSSQSENLPDQVSOREGIOAEQIVIKI 74
61 TDGQVYTSBGHDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGYYIKVQDKY 120
75 TDGQVYTSBGHDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGYYIKVQDKY 134
121 VYVLLKDAHADNVRTK 136
135 VYVLLKDAHADNVRTK 150

RESULT 3

AA81708 standard; Protein: 484 AA.

AA81708;

02-JUN-2000 (first entry)

Streptococcus pneumoniae protein sequence ID128.

Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
kidney disease; diabetes; immunosuppressive disorder; otitis media;
pneumococcal septicaemia; sinusitis; meningitis; therapy.

Streptococcus pneumoniae.

W0200006738-A2.

10-FEB-2000.

27-JUL-1999; 99MO-GB02452.

27-JUL-1998; 98GB-0016336.

19-MAR-1999; 99US-0125329.

XX (MICH-) MICROBIAL TECHNIQS LTD.
XX Le Page RW, Wells JM, Hanniffy SB, Hansbro PM;
XX WPI: 2000-195301/17.
XX N-PSDB: AA291804.

XX Streptococcal proteins and polynucleotides useful for diagnosis,
XX treatment and prophylaxis of bacterial infections
XX
XX Claim 1: Page 39; 76pp; English.

XX This sequence represents a Streptococcus pneumoniae protein of the
XX invention. The proteins (or their homologues, derivatives and/or
XX fragments) are useful as immunogens or antigens. Immunogenic or antigenic
XX compositions comprising the proteins are useful as vaccines and also in
XX diagnostic assays. The sequences are useful for the detection or
XX diagnosis of S. pneumoniae infection, by contacting a sample to be tested
XX with them. Agents capable of antagonising, inhibiting or interfering with
XX the function or expression of the protein or polypeptide are useful in
XX medical compositions in the treatment or prophylaxis of S. pneumoniae
XX infection. As the sequences can be used to treat S. pneumoniae infection,
XX they can be used to treat bacterial pneumonia, which has high rates in
XX young children, the elderly, and in patients with predisposing conditions
XX such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
XX or with immunosuppressive disorders, especially AIDS. They can also be
XX used to treat pneumococcal septicaemia, otitis media, sinusitis, and
XX meningitis.

XX Sequence 484 AA:

Query Match 100.0%; Score 711; DB 21; Length 484;
Best Local Similarity 100.0%; Pred. No. 8.3e-70;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 IYSLSCATALNQHRSQEKDNRRVSYVDGSSQSKSENLTPDOVSQKKGIOAEQIVIRI 60
15 IYSLSCATALNQHRSQEKDNRRVSYVDGSSQSKSENLTPDOVSQKKGIOAEQIVIRI 74
61 TPOGVTSSGDRHYNGKVPYDALFSEELMKDPYQKADIVNEVGKGIIRYDGRY 120
75 TdggvYtsgdihyngkvpYdalFseelImkdpYqKadivnevggYlIkvdgry 134
QY 121 YYTLKDAHADNVRRK 136
135 YYTLKDAHADNVRRK 150
DB 135 YYTLKDAHADNVRRK 150

RESULT 4
ID AAB1538 standard; Protein; 485 AA.

XX AAB1538;
XX
XX 24-MAY-2000 (first entry)
XX Streptococcus pneumoniae type 4 protein sequence #38.
XX DE Streptococcus pneumoniae; vaccine; screening; protein antigen;
XX KW Streptococcus pneumoniae; vaccine; screening; protein antigen;
XX anti-bacterial; anti-inflammatory; meningitis; infection; diagnosis;
XX pneumococcal disease.
XX Streptococcus pneumoniae.
XX OS
XX PN WO200006737-A2.
XX 10-FEB-2000.
XX PD 27-JUL-1999; 99WO-GB02451.
XX PF 27-JUL-1998; 98GB-0016337.
XX PR

PR 19-MAR-1999; 99US-0125164.
XX (MICH-) MICROBIAL TECHNIQS LTD.
XX
XX Gilbert CFG, Hansbro PM;
XX WPI: 2000-195300/17.

XX New Streptococcal protein, useful as a vaccine, for diagnosis of
XX pneumococcal diseases and for screening agents capable of antagonizing
XX or inhibiting expression of the protein
XX
XX Claim 1: Page 76; 108pp; English.

XX AAB1501 to AAB1679 represent specifically claimed protein sequences
XX isolated from Streptococcus pneumoniae. AAB1501 to AAB1590 represent
XX specifically claimed nucleotide sequences isolated from S. pneumoniae.
XX The sequences have antibacterial and anti-inflammatory properties.
XX The protein sequences, and fragments of them, are useful as immunogens
XX and/or antigens. The nucleotide sequences can be used in vaccines and in
XX diagnostic assays. The proteins and nucleotides can be useful for the
XX detection and diagnosis of S. pneumoniae. The protein sequences are also
XX useful for screening an agent capable of antagonising, inhibiting or
XX interfering with the function or expression of the proteins in which the
XX agent is useful for treatment or prophylaxis of S. pneumoniae infection
XX and meningitis. AAB15591 to AAB15614 represent primers used in the
XX exemplification of the present invention.

XX Sequence 485 AA:

Query Match 100.0%; Score 711; DB 21; Length 485;
Best Local Similarity 100.0%; Pred. No. 8.3e-70;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IYSLSCATALNQHRSQEKDNRRVSYVDGSSQSKSENLTPDOVSQKKGIOAEQIVIRI 60
15 IYSLSCATALNQHRSQEKDNRRVSYVDGSSQSKSENLTPDOVSQKKGIOAEQIVIRI 74
61 TPOGVTSSGDRHYNGKVPYDALFSEELMKDPYQKADIVNEVGKGIIRYDGRY 120
75 TdggvYtsgdihyngkvpYdalFseelImkdpYqKadivnevggYlIkvdgry 134
QY 121 YYTLKDAHADNVRRK 136
135 YYTLKDAHADNVRRK 150
DB 135 YYTLKDAHADNVRRK 150

RESULT 5
ID AAB12724 standard; Protein; 509 AA.

XX AAB12724;
XX
XX 21-NOV-2000 (first entry)
XX Streptococcus pneumoniae L-BVH-3-AD protein antigen SEQ ID NO:57.
XX DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteremia;
XX otitis media; pneumoniae; immunisation; bactericidal.
XX Streptococcus pneumoniae.
XX OS
XX PN WO200039299-A2.
XX 06-JUL-2000.
XX PD 20-DEC-1999; 99WO-CA01218.
XX PF 23-DEC-1998; 98US-0113800.
XX PR (BIOC-) BIOCHEM PHARMA INC.

KX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 PI WPI: 2000-452397/39.
 KX Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteremia and/or pneumonia
 KX Claim 18; Fig 22; 106pp; English.
 KX The present invention describes nucleic acids (I) encoding protein
 KX antigens (II) from Streptococcus pneumoniae. The protein antigens
 KX have bactericidal activity. The nucleic acids, encoding the protein
 KX antigens, may be used for the recombinant production of the proteins
 KX they encode. The protein antigens may then be used as vaccines for the
 KX prevention and treatment of Streptococcal infections in mammals
 KX (especially humans) which result in, e.g. meningitis, otitis media,
 KX bacteremia and/or pneumonia. The present sequence represents the
 KX S. pneumoniae L-BVH-3 AD protein antigen.

Sequence 509 AA;

Query Match 100.0%; Score 711; DB 21; Length 509;

Best Local Similarity 100.0%; Pred. No. 8.9e-70; Mismatches 0; Gaps 0;

Matches 136; Conservative 0; Indels 0; Gaps 0;

1 IVSLICAYALNORSEKNDNRVSYDGSQSSOKSENTLPDOVSQREGIOAQIVIKI 60

15 IIVSLICAYALNORSEKNDNRVSYDGSQSSOKSENTLPDOVSQREGIOAQIVIKI 74

61 TDGQVYVSHGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEKGYIIRVDKRY 120

75 TDGQVYVSHGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEKGYIIRVDKRY 134

121 YVYIKDAAHADNVTK 136

135 YVYIKDAAHADNVTK 150

RESULT 6

AA12715 standard; Protein; 1039 AA.

AA12715;

21-NOV-2000 (first entry)

Streptococcus pneumoniae BVH-3 protein antigen SEQ ID NO:2.

Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

Streptococcus pneumoniae; infection; diagnosis; meningitis; bacteremia;

otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.

MO200039299-A2.

06-JUL-2000.

20-DEC-1999; 99MO-CA01218.

23-DEC-1998; 98US-0113800.

(BIOC-) BIOCHEM PHARMA INC.

Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

WPI: 2000-452397/39.

N-PSDB; AAA65730.

Streptococcal antigens useful for vaccinating against e.g. meningitis,
 otitis media, bacteremia and/or pneumonia

KX Claim 18; Fig 2; 106pp; English.
 PS
 KX

CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteremia and/or pneumonia. The present sequence represents the
 CC S. pneumoniae BVH-3 protein antigen.

Sequence 1039 AA;

Query Match 100.0%; Score 711; DB 21; Length 1039;

Best Local Similarity 100.0%; Pred. No. 2.4e-69; Mismatches 0; Gaps 0;

Matches 136; Conservative 0; Indels 0; Gaps 0;

1 IVSLICAYALNORSEKNDNRVSYDGSQSSOKSENTLPDOVSQREGIOAQIVIKI 60

15 IIVSLICAYALNORSEKNDNRVSYDGSQSSOKSENTLPDOVSQREGIOAQIVIKI 74

61 TDGQVYVSHGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEKGYIIRVDKRY 120

75 TDGQVYVSHGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEKGYIIRVDKRY 134

121 YVYIKDAAHADNVTK 136

135 YVYIKDAAHADNVTK 150

RESULT 7

AA12726 standard; Protein; 205 AA.

AA12726;

21-NOV-2000 (first entry)

Streptococcus pneumoniae BVH-3C protein antigen SEQ ID NO:59.

Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

Streptococcus pneumoniae; infection; diagnosis; meningitis; bacteremia;

otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.

MO200039299-A2.

06-JUL-2000.

20-DEC-1999; 99MO-CA01218.

23-DEC-1998; 98US-0113800.

(BIOC-) BIOCHEM PHARMA INC.

Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

WPI: 2000-452397/39.

Streptococcal antigens useful for vaccinating against e.g. meningitis,
 otitis media, bacteremia and/or pneumonia

Streptococcal antigens useful for vaccinating against e.g. meningitis,
 otitis media, bacteremia and/or pneumonia

Claim 18; Fig 24; 106pp; English.

The present invention describes nucleic acids (I) encoding protein
 antigens (II) from Streptococcus pneumoniae. The protein antigens
 have bactericidal activity. The nucleic acids, encoding the proteins
 they encode, may be used for the recombinant production of the proteins
 they encode. The protein antigens may then be used as vaccines for the

prevention and treatment of Streptococcal infections in mammals (especially humans) which result in, e.g. meningitis, otitis media, bacteraemia and/or pneumonia. The present sequence represents the S. pneumoniae BVH-3C protein antigen.

Sequence 205 AA;

Query Match 96.6%; Score 687; DB 21; Length 205;
Best Local Similarity 100.0%; Pred. No. 1.1e-67;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CAYALNQHRSQENKNNRRSYVDGSSQSKSENLTPDQVSQKKGIAQAEQIVIKITDQGYV 66
1 cayalnghtsrgenknrrsyvdsqsgsksenltpdqvsqkkgiaeqivikltdqgyv 60
QY 67 TSHGDHYHYNGKVPYDALFSEELMKDPYOLKADIVNEVGKGIIRKVDGKRYYYLKD 126
61 tshgdhyhyngkvpdyalfseellmkdpnyqlkdadivnevsgyylkvdgkryyyylkd 120
QY 127 AAADNVRRK 136
121 aaadnvrrk 130
Db

RESULT 8

AAB12723 8
ID AAB12723 standard; Protein; 489 AA.

AAB12723;

DT 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae BVH-3AD protein antigen SEQ ID NO:56.

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

KM otitis media; pneumonia; immunisation; bactericidal.

OS Streptococcus pneumoniae.

XX MO200039299-A2.

06-JUL-2000.

20-DEC-1999; 99WO-CA01218.

23-DEC-1996; 98US-0113800.

(BIOC-) BIOCHEM PHARMA INC.

Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

WPI; 2000-452397/39.

PT Streptococcal antigens useful for vaccinating against e.g. meningitis,

OTitis media, bacteraemia and/or pneumonia.

Claim 18; Fig 21; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the proteins
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae BVH-3AD protein antigen.

Sequence 489 AA;

Query Match 96.6%; Score 687; DB 21; Length 489;
Best Local Similarity 100.0%; Pred. No. 3.7e-67;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CAYALNQHRSQENKNNRRSYVDGSSQSKSENLTPDQVSQKKGIAQAEQIVIKITDQGYV 66
1 cayalnghtsrgenknrrsyvdsqsgsksenltpdqvsqkkgiaeqivikltdqgyv 60
QY 67 TSHGDHYHYNGKVPYDALFSEELMKDPYOLKADIVNEVGKGIIRKVDGKRYYYLKD 126
61 tshgdhyhyngkvpdyalfseellmkdpnyqlkdadivnevsgyylkvdgkryyyylkd 120
QY 127 AAADNVRRK 136
121 aaadnvrrk 130
Db

RESULT 9

AAB12744 9
ID AAB12744 standard; Protein; 780 AA.

AAB12744;

DT 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae NEM15 protein antigen SEQ ID NO:78.

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

KM otitis media; pneumonia; immunisation; bactericidal.

OS Streptococcus pneumoniae.

XX MO200039299-A2.

06-JUL-2000.

20-DEC-1999; 99WO-CA01218.

23-DEC-1996; 98US-0113800.

(BIOC-) BIOCHEM PHARMA INC.

Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

WPI; 2000-452397/39.

PT Streptococcal antigens useful for vaccinating against e.g. meningitis,

OTitis media, bacteraemia and/or pneumonia.

Claim 18; Fig 43; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the proteins
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae NEM15 protein antigen.

Sequence 780 AA;

Query Match 96.6%; Score 687; DB 21; Length 780;
Best Local Similarity 100.0%; Pred. No. 7.1e-67;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CAYALNQHRSQENKNNRRSYVDGSSQSKSENLTPDQVSQKKGIAQAEQIVIKITDQGYV 66
1 cayalnghtsrgenknrrsyvdsqsgsksenltpdqvsqkkgiaeqivikltdqgyv 60
Db

QY 67 TSHGHHYHYNGKVPYDALFSEELLMKDPNTQKADIVNEKGGITIKVDGKYYVYLKD 126
DB 61 tshgddhyhyngkvpdydalfeelllmkdpnyqjlkdadivnevkggyllkvdgkyyvylkd 120
QY 127 AAHADNVTRK 136
DB 121 aahadnvtrk 130

RESULT 10

AAH12721

ID AAB12721 standard; Protein; 840 AA.

AC AAB12721;

DT 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae Sp63 BVH-3 protein antigen SEQ ID NO:16.

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

KW otitis media; pneumoniae; immunisation; bactericidal.

OS Streptococcus pneumoniae.

MO200039299-A2.

06-JUL-2000.

PF 20-DEC-1999; 99MO-CA01218.

PR 23-DEC-1998; 98US-0113800.

PA (BIOC-) BIOCHEM PHARMA INC.

PI Hamel J, Brodeur BR, Pleneau I, Martin D, Rioux C, Charland N;

DR WPI: 2000-452397/39.

N-PSDB; AAA65738.

PT Streptococcal antigens useful for vaccinating against e.g. meningitis,

PT otitis media, bacteraemia and/or pneumonia

PS Claim 18; Fig 19; 106pp; English.

3C The present invention describes nucleic acids (I) encoding protein
3C antigens (II) from Streptococcus pneumoniae. The protein antigens
3C have bactericidal activity. The nucleic acids, encoding the protein
3C antigens, may be used for the recombinant production of the proteins
3C they encode. The protein antigens may then be used as vaccines for the
3C prevention and treatment of Streptococcal infections in mammals
3C (especially humans) which result in, e.g. meningitis, otitis media,
3C bacteraemia and/or pneumonia. The present sequence represents the
3C S. pneumoniae Sp63 BVH-3 protein antigen.

Sequence 840 AA;

Query Match 96.6%; Score 687; DB 21; Length 840;
Best Local Similarity 100.0%; Pred. No. 7.9e-67;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CAYALNHRSGENKDNRRVSYVDSQSSQKSENLPDQVSGREGIOAEQIVIKITDGGTV 66

DB 1 cayalnhrsgenkdnnrvsyvdsqsgsqskenltpdqvsqgqiaeqivlkitdggv 60

QY 67 TSHGHHYHYNGKVPYDALFSEELLMKDPNTQKADIVNEKGGITIKVDGKYYVYLKD 126

DB 61 tshgddhyhyngkvpdydalfeelllmkdpnyqjlkdadivnevkggyllkvdgkyyvylkd 120

QY 127 AAHADNVTRK 136

DB 121 aahadnvtrk 130

RESULT 11
AAB12722
ID AAB12722 standard; Protein; 1019 AA.
XX
AC AAB12722;

DT 21-NOV-2000 (first entry).

DE Streptococcus pneumoniae BVH-3M protein antigen SEQ ID NO:55.

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

KW otitis media; pneumoniae; immunisation; bactericidal.

OS Streptococcus pneumoniae.

MO200039299-A2.

06-JUL-2000.

PF 20-DEC-1999; 99MO-CA01218.

PR 23-DEC-1998; 98US-0113800.

PA (BIOC-) BIOCHEM PHARMA INC.

PI Hamel J, Brodeur BR, Pleneau I, Martin D, Rioux C, Charland N;

DR WPI: 2000-452397/39.

N-PSDB; AAA65738.

PT Streptococcal antigens useful for vaccinating against e.g. meningitis,

PT otitis media, bacteraemia and/or pneumonia

PS Claim 18; Fig 20; 106pp; English.

3C The present invention describes nucleic acids (I) encoding protein
3C antigens (II) from Streptococcus pneumoniae. The protein antigens
3C have bactericidal activity. The nucleic acids, encoding the proteins
3C antigens, may be used for the recombinant production of the proteins
3C they encode. The protein antigens may then be used as vaccines for the
3C prevention and treatment of Streptococcal infections in mammals
3C (especially humans) which result in, e.g. meningitis, otitis media,
3C bacteraemia and/or pneumonia. The present sequence represents the
3C S. pneumoniae BVH-3M protein antigen.

Sequence 1019 AA;

Query Match 96.6%; Score 687; DB 21; Length 1019;
Best Local Similarity 100.0%; Pred. No. 1e-66;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CAYALNHRSGENKDNRRVSYVDSQSSQKSENLPDQVSGREGIOAEQIVIKITDGGTV 66

DB 1 cayalnhrsgenkdnnrvsyvdsqsgsqskenltpdqvsqgqiaeqivlkitdggv 60

QY 67 TSHGHHYHYNGKVPYDALFSEELLMKDPNTQKADIVNEKGGITIKVDGKYYVYLKD 126

DB 61 tshgddhyhyngkvpdydalfeelllmkdpnyqjlkdadivnevkggyllkvdgkyyvylkd 120

QY 127 AAHADNVTRK 136

DB 121 aahadnvtrk 130

RESULT 12

AAB12748

ID AAB12748 standard; Protein; 1019 AA.

AC AAB12748;

21-NOV-2000 (first entry)
Streptococcus pneumoniae strain WU2 BVH-3 protein antigen.
Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
otitis media; pneumonia; immunisation; bactericidal.
Streptococcus pneumoniae.
MO200039299-A2.
06-JUL-2000.
20-DEC-1999; 99MO-CA01218.
23-DEC-1998; 98US-0113800.
(BIOC-) BIOCHEM PHARMA INC.
Hamel J, Brodeur BR, Plneau I, Martin D, Rioux C, Charland N;
MPI; 2000-452397/39.
Streptococcal antigens useful for vaccinating against e.g. meningitis,
otitis media, bacteraemia and/or pneumonia -
Disclosure; Fig 11; 106pp; English.
The present invention describes nucleic acids (I) encoding protein
antigens (II) from Streptococcus pneumoniae. The protein antigens
have bactericidal activity. The nucleic acids, encoding the proteins
antigens, may be used for the recombinant production of the proteins
they encode. The protein antigens may then be used as vaccines for the
prevention and treatment of Streptococcal infections in mammals
(especially humans) which result in, e.g. meningitis, otitis media,
bacteraemia and/or pneumonia. The present sequence represents a
S. pneumoniae BVH-3 protein antigen, from the present invention.
Sequence 1019 AA;

Query Match 96.6%; Score 687; DB 21; Length 1019;
Best Local Similarity 100.0%; Pred. No. 1e-66;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

7 CAYALNHRSGENKDNKRVSYVDGSSQSKSENTLPDQVSGKEIOAEQIYIKITDGYV 66
1 cayalnhrsgenkdnrsvyvdgsgsqsksentlpdqvsgkeiqvikiitdgyv 60
67 TSHGDHYHYNGKVPYDALFSEELMKDPNTQLKADIVNEVGKGIITKVDGKTYVYLKD 126
61 tshgdhyhyngkvpdyalfseellmkdpnyqlkadivnevgkgyllkvdgktyvylkd 120
127 AAHADNVRTK 136
121 aahadnvr tk 130

RESULT 13
AA012749
AA012749 standard; Protein; 1019 AA.

21-NOV-2000 (first entry)
Streptococcus pneumoniae strain RX1 BVH-3 protein antigen.
Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.
MO200039299-A2.
06-JUL-2000.
20-DEC-1999; 99MO-CA01218.
23-DEC-1998; 98US-0113800.
(BIOC-) BIOCHEM PHARMA INC.
Hamel J, Brodeur BR, Plneau I, Martin D, Rioux C, Charland N;
MPI; 2000-452397/39.
Streptococcal antigens useful for vaccinating against e.g. meningitis,
otitis media, bacteraemia and/or pneumonia -
Disclosure; Fig 11; 106pp; English.
The present invention describes nucleic acids (I) encoding protein
antigens (II) from Streptococcus pneumoniae. The protein antigens
have bactericidal activity. The nucleic acids, encoding the proteins
antigens, may be used for the recombinant production of the proteins
they encode. The protein antigens may then be used as vaccines for the
prevention and treatment of Streptococcal infections in mammals
(especially humans) which result in, e.g. meningitis, otitis media,
bacteraemia and/or pneumonia. The present sequence represents a
S. pneumoniae BVH-3 protein antigen, from the present invention.
Sequence 1019 AA;

Query Match 96.6%; Score 687; DB 21; Length 1019;
Best Local Similarity 100.0%; Pred. No. 1e-66;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

7 CAYALNHRSGENKDNKRVSYVDGSSQSKSENTLPDQVSGKEIOAEQIYIKITDGYV 66
1 cayalnhrsgenkdnrsvyvdgsgsqsksentlpdqvsgkeiqvikiitdgyv 60
67 TSHGDHYHYNGKVPYDALFSEELMKDPNTQLKADIVNEVGKGIITKVDGKTYVYLKD 126
61 tshgdhyhyngkvpdyalfseellmkdpnyqlkadivnevgkgyllkvdgktyvylkd 120
127 AAHADNVRTK 136
121 aahadnvr tk 130

RESULT 14
AA012750
AA012750 standard; Protein; 1019 AA.

21-NOV-2000 (first entry)
Streptococcus pneumoniae strain JNR7/87 BVH-3 protein antigen.
Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.
MO200039299-A2.
06-JUL-2000.
20-DEC-1999; 99MO-CA01218.

PR 23-DEC-1998; 9805-0113800.
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 OR WPI; 2000-452397/39.
 XX
 XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 XX
 XX Disclosure; Fig 11; 106pp; English.
 PS
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins.
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-3 protein antigen, from the present invention.
 XX
 SO Sequence 1019 AA;

Query Match 96.6%; Score 687; DB 21; Length 1019;
 Best Local Similarity 100.0%; Pred. No. 1e-66;
 Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 7 CAYALNQHRSQENKNNRYSYVDSGSSQKSENLPDQVSQKGIQAEQIVYKITDQGYV 66
 |||||||
 Db 1 cayalnqhnsqenkdnnrsvydgsgsqksenltpdqvsqkqiaeqivlkitdqgyv 60
 67 TSHGDHYHYNGKVPYDALFSEELMKDPYOLKADIVNEVGGYIIKVDGKYYVYI 126
 |||||||
 Db 61 tshgdhyhyngkvpdydalalseellmkdpnyqlkdadivnevggyiikvdkgyyyvylnd 120
 2Y 127 AAHADNVRTK 136
 |||||||
 Db 121 aahadvrtk 130

RESULT 15
 AAB12751
 ID AAB12751 standard; Protein; 1019 AA.
 XX
 XX AAB12751;
 XX
 XX 21-NOV-2000 (first entry)
 XX
 XX Streptococcus pneumoniae strain SP64 BVH-3 protein antigen.
 XX
 XX Streptococcus pneumoniae: BVH-3; BVH-11; BVH-28; antigen; vaccine;
 XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 XX otitis media; pneumonia; immunisation; bactericidal.
 XX
 XX Streptococcus pneumoniae.
 XX
 XX MO200039299-AA2.
 XX
 XX 06-JUL-2000.
 XX
 XX 20-DEC-1999; 99WO-CA01218.
 XX
 XX 23-DEC-1998; 9805-0113800.
 XX
 XX (BIOC-) BIOCHEM PHARMA INC.
 XX
 XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX WPI; 2000-452397/39.
 XX
 XX

PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 XX
 XX Disclosure; Fig 11; 106pp; English.
 PS
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-3 protein antigen, from the present invention.
 XX
 SO Sequence 1019 AA;

Query Match 96.6%; Score 687; DB 21; Length 1019;
 Best Local Similarity 100.0%; Pred. No. 1e-66;
 Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CAYALNQHRSQENKNNRYSYVDSGSSQKSENLPDQVSQKGIQAEQIVYKITDQGYV 66
 |||||||
 Db 1 cayalnqhnsqenkdnnrsvydgsgsqksenltpdqvsqkqiaeqivlkitdqgyv 60
 QY 67 TSHGDHYHYNGKVPYDALFSEELMKDPYOLKADIVNEVGGYIIKVDGKYYVYI 126
 |||||||
 Db 61 tshgdhyhyngkvpdydalalseellmkdpnyqlkdadivnevggyiikvdkgyyyvylnd 120
 QY 127 AAHADNVRTK 136
 |||||||
 Db 121 aahadvrtk 130

Search completed: September 26, 2001, 22:30:30
 Job time: 3209 sec

ALL INFORMATION CONTAINED
HEREIN IS UNCLASSIFIED
DATE 10-05-2001 BY 60324

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

Protein - protein search, using sw model

Run on: September 26, 2001, 22:33:45 ; Search time 73.11 Seconds
(without alignments)
246.115 Million cell updates/sec

Title: US-09-471-255-2_COPY_15_150

Correct score: 711
Sequence: 1 IVSLSLCAVALNQHRSQENK.....DKYTYVLKDAHADNVRK 136

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SPTREMBL_16:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mmc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_unclassified:*
- 13: sp_vertebrate:*
- 14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 423 | 59.5 | 822 | 2 | Q9ZHG7 streptococ |
| 2 | 103 | 14.5 | 1089 | 10 | Q9FN97 arabidopsi |
| 3 | 88 | 12.4 | 198 | 2 | 053037 haemophil |
| 4 | 87.5 | 12.3 | 736 | 13 | 093583 rana ridib |
| 5 | 85.5 | 12.0 | 721 | 2 | 09PR88 ureaplasma |
| 6 | 84.5 | 11.9 | 1877 | 2 | Q9XKX1 plasmodium |
| 7 | 83 | 11.7 | 814 | 2 | Q9PHD9 xylella fas |
| 8 | 82 | 11.5 | 1946 | 2 | Q48545 lactobacilli |
| 9 | 81.5 | 11.5 | 703 | 5 | Q46944 aplysia cal |
| 10 | 81 | 11.4 | 712 | 5 | Q16971 aplysia cal |
| 11 | 79.5 | 11.2 | 567 | 2 | Q9ZM97 helicobacte |
| 12 | 79.5 | 11.2 | 2573 | 5 | Q96185 plasmodium |
| 13 | 78 | 11.0 | 264 | 2 | Q95357 mycoplasma |
| 14 | 78 | 11.0 | 438 | 2 | Q46298 clostridium |
| 15 | 78 | 11.0 | 642 | 2 | Q9PHM6 campylobac |
| 16 | 77.5 | 10.9 | 206 | 3 | Q74190 kluyveromyc |
| 17 | 77.5 | 10.9 | 857 | 10 | Q9T0C8 arabidopsi |
| 18 | 77 | 10.8 | 776 | 5 | Q15793 p. strain dd |
| 19 | 77 | 10.8 | 1766 | 5 | Q25668 plasmodium |

| | | | | | |
|----|------|------|------|----|---------------------|
| 20 | 77 | 10.8 | 1785 | 5 | Q25685 plasmodium |
| 21 | 76.5 | 10.8 | 490 | 2 | Q9RMV1 bacillus an |
| 22 | 76.5 | 10.8 | 1799 | 5 | Q18220 caenorhabd |
| 23 | 76.5 | 10.8 | 1802 | 5 | Q18219 caenorhabd |
| 24 | 76 | 10.7 | 608 | 10 | Q9FFG6 arabidopsi |
| 25 | 76 | 10.7 | 875 | 5 | Q9P7S5 schizosach |
| 26 | 76 | 10.7 | 2421 | 5 | Q9VM88 drosophila |
| 27 | 75.5 | 10.6 | 312 | 2 | Q50603 staphylococ |
| 28 | 75.5 | 10.6 | 332 | 9 | Q80066 bacterioph |
| 29 | 75.5 | 10.6 | 472 | 5 | Q9XOB6 drosophila |
| 30 | 75 | 10.5 | 214 | 5 | Q18307 caenorhabd |
| 31 | 75 | 10.5 | 1060 | 5 | Q16926 aedes aegypt |
| 32 | 74.5 | 10.5 | 388 | 1 | Q28907 archaeoglob |
| 33 | 74.5 | 10.5 | 753 | 6 | Q9GLR1 bos taurus |
| 34 | 74 | 10.4 | 710 | 4 | Q9UM02 homo sapien |
| 35 | 74 | 10.4 | 802 | 5 | Q9V773 drosophila |
| 36 | 73.5 | 10.3 | 472 | 5 | Q9GR08 plasmodium |
| 37 | 73.5 | 10.3 | 1007 | 5 | Q96848 dictyostell |
| 38 | 73 | 10.3 | 282 | 2 | Q50852 borrelia bu |
| 39 | 73 | 10.3 | 506 | 4 | Q14898 homo sapien |
| 40 | 73 | 10.3 | 556 | 2 | Q9XW3 porphyromon |
| 41 | 73 | 10.3 | 556 | 2 | Q9R072 porphyromon |
| 42 | 73 | 10.3 | 710 | 11 | Q70196 rattus norv |
| 43 | 73 | 10.3 | 710 | 11 | Q9QUR6 mus musculu |
| 44 | 73 | 10.3 | 1590 | 2 | Q59983 streptococ |
| 45 | 72.5 | 10.2 | 385 | 14 | Q9IFJ1 helicoverpa |

ALIGNMENTS

| RESULT | 1 | PREDIMINARY | PRT | 822 AA. |
|-----------------------|--|-------------|-----|---------|
| Q9ZHG7 | Q9ZHG7 | | | |
| AC | Q9ZHG7 | | | |
| DT | 01-MAY-1999 (TREMBLrel. 10, Created) | | | |
| DT | 01-MAY-1999 (TREMBLrel. 10, Last sequence update) | | | |
| DT | 01-MAY-2000 (TREMBLrel. 13, Last annotation update) | | | |
| DE | HYPOTHETICAL 92.4 KDA PROTEIN. | | | |
| OS | Streptococcus agalactiae. | | | |
| OC | Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; | | | |
| OC | Streptococcus. | | | |
| OX | NCBI_TaxID-1311; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN-R268; | | | |
| RX | MEDLINE-99115568; Pubmed-9916102; | | | |
| RA | Spellierberg B., Rozdzinski E., Martin S., Weber-Heymann J., | | | |
| RA | Schmitzler N., Luetticken R., Podbielski A.; | | | |
| RT | "Lmb, a protein with similarities to the Lrai adhesin family, mediates | | | |
| RT | attachment of Streptococcus agalactiae to human laminin." | | | |
| RL | Infect. Immun. 67:871-878(1999). | | | |
| DR | EMBL; AF062533; AAD13797.1; | | | |
| KW | Hypothetical protein. | | | |
| SO | SEQUENCE 822 AA; 92385 MW; 80E4EDF313481F98 CRC64; | | | |
| Query Match | 59.5%; Score 423; DB 2; Length 822; | | | |
| Best Local Similarity | 55.1%; Pred. No. 4, 2e-29; | | | |
| Matches | 76; Conservative 26; Mismatches 34; Indels 2; Gaps 1; | | | |
| QY | 1 IVSLSLCAVALNQHRSQENKNNRVSYYDGSQSSQKS--ENTTPQVQSKGIOAEQIV 58 | | | |
| DB | 15 LLAFTIGSYOUGKHMGLATKMDQIAYTIDDSGKVKAKTKNTMDQISAESTISAEQIV 74 | | | |
| QY | 59 KITDGGYTSQSDHYHYNGKVPYDALFSEELMADPNYQLKADIVAEVKGQYIKYNG 118 | | | |
| DB | 75 KITDGGYTSQSDHYHYNGKVPYDALFSEELMADPNYHFGQSVINELIDGYIYKNG 134 | | | |
| QY | 119 KYTYVLKDAHADNVRK 136 | | | |
| DB | 135 MYTYVLKSGSKKNIRTK 152 | | | |

```

RESULT 2
ID 09FN97 PRELIMINARY; PRT: 1089 AA.
C 09FN97;
T 01-MAR-2001 (TREMblrel. 16, Created)
T 01-MAR-2001 (TREMblrel. 16, Last sequence update)
T 01-MAR-2001 (TREMblrel. 16, Last annotation update)
E TRANSPOSON PROTEIN-LIKE.
S Arabidopsis thaliana (Mouse-ear cross).
S Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
C Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
C Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
[1]
P SEQUENCE FROM N.A.
C STRAIN-COLOMBIA;
X MEDLINE-98069011; PubMed-9405937;
A Kotani H., Nakamura T., Sato S., Kaneko T., Asamizu E., Miyajima N.,
A Tabata S.;
T "Structural analysis of Arabidopsis thaliana chromosome 5. II.
T Sequence features of the regions of 1,044,062 bp covered by thirteen
T physically assigned pl clones."
L DNA Res. 4:291-300(1997).
P EMBL: AB006705; BAB09502.1;
F SEQUENCE 1089 AA; 126629 MW; 17664F2489EB14E CRC64;

Query Match 14.5%; Score 103; DB 10; Length 1089;
Best Local Similarity 23.7%; Pred. No. 0.76;
Matches 31; Conservative 25; Mismatches 73; Indels 2; Gaps 2;

Y 5 SLCAVALNHRSGENKNNRVSYYDGSQSSQSKSENTLPDQVSQKEGIAEQIVIKITDQ 64
b 792 TFCYTYGDBHKTNRAGDR HYDGNQEDHEHDYDPDIFSQAGRSKSEWLQDKD 850
Y 65 YVSHGSHHYNKNKYVDALFSEELMKDPNYOLKADIVNEVK-GGYIIKVDGKYYV 123
b 851 YHMHRYTLRNCDDLRFERFLFDESLIANPGISKDLNELREKQSSWLKKNYSPEW 910
Y 124 LKDAAHADNVR 134
b 911 LLSIVHGPMVK 921

ESTAT 3
ID 053037 PRELIMINARY; PRT: 198 AA.
C 053037;
T 01-JUN-1998 (TREMblrel. 06, Created)
T 01-JUN-1998 (TREMblrel. 06, Last sequence update)
T 01-NOV-1998 (TREMblrel. 08, Last annotation update)
F PUTATIVE HAEMOCIN PROCESSING PROTEIN.
P HMC.
S Haemophilus influenzae.
S Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
S Haemophilus.
NCBI_TaxID=727;
[1]
P SEQUENCE FROM N.A.
C STRAIN-ELA;
X MEDLINE-97197533; PubMed-9045829;
A Murley Y.M., Edling T.D., Pozsgay J.M., Lipuma J.J.;
T "Cloning and characterization of the haemocin immunity gene of
T Haemophilus influenzae."
J. Bacteriol. 179:1684-1689(1997).
[2]
P SEQUENCE FROM N.A.
C STRAIN-ELA;
A Murley Y.M., Edling T.D., Pozsgay J.M., Lipuma J.J.;
T Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
[3]
P SEQUENCE FROM N.A.

```

```

RC STRAIN-ELA;
RA Murley Y.M., Edling T.D., Pozsgay J.M., Lipuma J.J.;
RT Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: U68399; AAC46223.1;
SO SEQUENCE 198 AA; 22978 MW; 4C3F3F36A8723C95 CRC64;

Query Match 12.4%; Score 88; DB 2; Length 198;
Best Local Similarity 25.9%; Pred. No. 2;
Matches 39; Conservative 28; Mismatches 39; Indels 42; Gaps 9;

Y 14 HRSQENK-----NNRVSYYDGSQSSQK-SENTLPDQVSQKEGIAEQIVIKITDQ 63
b 63 HISEINEDTILYKLNPNNEASFELARISKEFNINAGLTLTKELINIKPVI----- 117
Y 64 GYVS--HGSHYHNGKVPYDALFSEELMKDP--NYOLKAD-----IVNEVGKY 112
b 118 AYVNSLSNDFITING-----IFNKELLSDPAIGNYSLSKSPFERIMLNRDKKGI 171
Y 113 IIRYDGKYYVYLKDAH---ADNVRTK 136
b 172 L-----YLRSDKHLEFDINIKT 191

RESULT 4
ID 093583 PRELIMINARY; PRT: 736 AA.
AC 093583;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE PROHORMONE CONVERTASE 1 PRECURSOR.
OS Rana ridibunda (laughing frog) (Marsh frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.
NCBI_TaxID=8406;
[1]
P SEQUENCE FROM N.A.
RA Gangnon F., Jang S., Jegou S., Vleau D., Seidah N.G., Vaudry H.;
RT "Cloning of frog PC1."
J. Comp. Neurol. 0:0-0(1998).
DR EMBL: AF092904; AAC62483.1;
DR InterPro: IPR002884;
DR Pfam: PF00082; Peptidase_S8; 1.
DR Pfam: PF01483; P. 1.
DR PRINTS: PRO0723; SUBTILISIN.
DR PRODOM: PD000717; 1.
DR PROSITE: PS00136; SUBTILASE_ASP. 1.
DR PROSITE: PS00137; SUBTILASE_HIS. 1.
DR PROSITE: PS00138; SUBTILASE_SER. 1.
SO SEQUENCE 736 AA; 82965 MW; 84C00B44EB8C3CD7 CRC64;

Query Match 12.3%; Score 87.5; DB 13; Length 736;
Best Local Similarity 23.2%; Pred. No. 11;
Matches 36; Conservative 20; Mismatches 44; Indels 55; Gaps 5;

Y 5 SLCAVALNHRSGENK-----DNNRVSYYDGSQSSQK-----SENT 41
b 63 SLGTHFLFRHRDPRRSRSAPATIRKLVDNRSWAEOOYIQORRGYVANTDSEDLF 122
Y 42 PDQVS-----KEGIAEQIVIKITDQGYVTSHGSHYHY 76
b 123 NDKVPMKNQVTLNDRVNPPLPKLIDLVIVYVWKRGLTGKGSVTVLDDGLENNHTDIYNT 182
Y 77 NGKVPYDALFSEELMKDPNYOLKADIVNEVGK 111
b 183 DPEASVD--FNDNDKPPRY-----DITNENKKG 210

RESULT 5
ID 09PR88

```

ID 09P88 PRELIMINARY: PRT: 721 AA.
 AC 09P88;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE VIRULENCE ASSOCIATED PROTEIN-EXORIBONUCLEASES.
 3N VACB.
 3M Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
 3C Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 3X Mycoplasmataceae; Ureaplasma.
 NCBI_TaxID=134821;
 RN [1].
 RP SEQUENCE FROM N.A.
 ST STRAIN-SEROVAR 3;
 MEDLINE=20500219; PubMed=11048724;
 Glas J.I., Lefkowitz E.J., Glas J.S., Helner C.R., Chen E.Y.,
 Caswell G.H.;
 RA "The complete sequence of the mucosal pathogen Ureaplasma
 RT urealyticum."
 RL Nature 407:757-762(2000).
 DR EMBL; AE002105; AAF30462.1;
 DR InterPro: IPR001900;
 DR InterPro: IPR003029;
 DR Pfam: PF00575; S1: 1;
 DR Pfam: PF00773; RMB; 1;
 DR PROSITE: PS01175; RIBONUCLEASE II: 1;
 SQ SEQUENCE 721 AA; 83173 MW; 67EEFA9AA4EFD4AA CRC64;

Query Match: 12.0%; Score 85.5; DB 2; Length 721;
 Best Local Similarity: 25.7%; Pred. No. 16; Mismatches 55; Indels 29; Gaps 7;
 Matches 36; Conservative 20;
 QY 11 LNHRSQENKDNRR--VSYDGSQSSQSKSENLPDVSQKEGIAEQIYKIRDOGVTS 68
 DB 48 LIOENTIKLENNRLVIGLYDFEHE-----IKGI-----ITINSGDGFKE 92
 QY 69 HGDHYHY-NGKVPYALFSEEL---LAKDPYQAKDA---DIVNEK---GGYTIKY 116
 DB 93 DMEIERYNKKTLNGLKDKSVKFKLKEPKNNLDAAVIEIGHANDHYGVGPIITP 152
 117 DQYVYVYKDAHADNVK 136
 153 NGGYIFVDDPLEYLNINLK 172

RESULT 6
 ID 09XKWL PRELIMINARY: PRT: 1877 AA.
 AC 09XKWL;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE P-TYPE ATPASE.
 OS Plasmodium yoelii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxID=5861;
 RN [1].
 RP SEQUENCE FROM N.A.
 ST STRAIN-17XL STRAIN;
 Kimura M., Tanabe K., Krishna S., Tsudoi T., Saito-Ita A., Ocani S.,
 Ogura H.;
 RA "Characterization of a novel P-type ATPase in Plasmodium yoelii."
 RT Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 EMBL; AB016798; BAA77326.1;
 DR EMBL; AB016798; BAA77326.1;
 DR InterPro: IPR001757;
 DR Pfam: PF00122; E1-E2_ATPase; 1.
 DR PRINTS; PR00119; CATATPASE.
 DR PROSITE; PS00154; ATPASE_E1_E2; UNKNOWN_1.
 SQ SEQUENCE 1877 AA; 220196 MW; 16974C388B01746E CRC64;

Query Match

11.9%; Score 84.5; DB 5; Length 1877;

Best Local Similarity: 21.8%; Pred. No. 62;
 Matches 31; Conservative 35; Mismatches 63; Indels 13; Gaps 5;
 QY 1 IYSLCAVALNQH-RSQENKDNRRVSYDGSQSSQSKSENLPD-QVSKGEGIAEQIYI 58
 DB 848 INSTLEKLNKNNRDEONYNOTNKEIIDKSNSEKDFKSPNSICKSSQSSDILF 907
 QY 59 KITDGYTSHGDHYHYNGKV-----PYALFSEELMDPYNQAKDAIVNEVKG 111
 DB 908 QSYDGEYQKSEHPIKDEKITYTHQSSHFQAKDEIRDAEIR-KDEIRNDEE-- 964
 QY 112 YIIKYDGKYYVYIKDAHADNV 133
 DB 965 -IRKNENFNHNSNLSANSNNI 985

RESULT 7
 ID 09PHD9 PRELIMINARY: PRT: 814 AA.
 AC 09PHD9;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE DNA GYRASE SUBUNIT B.
 GN XFO005.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 NCBI_TaxID=23711;
 RN [1].
 RP SEQUENCE FROM N.A.
 ST STRAIN-9A5C;
 RC MEDLINE=20365717; PubMed=10910347;
 RX Simpson A.J.G., Reimach F.C., Arruda P., Abreu F.A., Acencio M.,
 Alvaranga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,
 Bueno M.R.P., Camargo A.C., Camargo L.E.A., Carraro D.M., Carrer H.,
 Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 Coutinho L.V., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
 Facchini A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 Gardner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.J.P.,
 Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.R., Kitajima J.P.,
 Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Matsukuma A.Y.,
 Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 Menck C.F.M., Miracca E.C., Miyaki C.Y., Montello-Vitorello C.B.,
 Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 Nhami A.Jr., Nobrega-F.G., Nunes L.R., Oliveira M.A.,
 de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pequeiro J.B.,
 Queglio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 da Rosa V.E.Jr., de Sa R.G., Santelli R.V., Savasaki H.E.,
 da Silva A.C.R., da Silva A.M., da Silva F.R., Silva H.A.Jr.,
 da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 de Souza A.P., Terenzi M.F., Trufi D., Tsai S.M., Tsuhako M.H.,
 Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 Zago M.A., Zatz M., Melandri J., Setubal J.C.;
 RA "The genome sequence of the plant pathogen Xylella fastidiosa."
 RT Nature 406:151-159(2000).
 CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 CC -1- OF DOUBLE-STRANDED DNA (BY SIMILARITY).
 CC -1- SIMILARITY: TO DNA TOPOISOMERASE II FAMILY.
 CC EMBL; AE003855; AAF82818.1;
 DR EMBL; AE003855; AAF82818.1;
 DR InterPro: IPR000410;
 DR InterPro: IPR001241;
 DR InterPro: IPR002288;
 DR InterPro: IPR002936;
 DR InterPro: IPR003594;
 DR Pfam: PF00204; DNA_TopoII; 2.
 DR Pfam: PF00986; DNA_GYRASEB_C; 1.
 DR Pfam: PF01751; Toprim; 1.

PRINTS: PR00418; TP12FAMILY
 PROSITE: PS00177; TOPOISOMERASE_II; 1
 SMART: SM00387; HATPase_C; 1
 ATP-binding; Isomerase; Topoisomerase
 SEQUENCE 814 AA; 90188 MW; 7C1EE231BC9A28A CRC64;

Query Match 11.7%; Score 83; DB 2; Length 814;
 Best Local Similarity 20.6%; Pred. No. 31;
 Matches 28; Conservative 21; Mismatches 45; Indels 42; Gaps 5;

27 YDGSQSSQKSNLTPDOVSGKEGIAQIYV-----KITDQGYTS--HG----- 70
 75 HVDGSVYSNDKRGIPVDHKEGSALEVITLVHAGSGEDDSKYSGGHGVGSVY 134
 71 -----DHYHYNGKVPDALFSEELMKDPYOLKADIYNEVKGYYIKVD 117
 135 NALSRLMLDIWRDGYH-----QGEYVLGEPOYPLKOLGV--SAKRGTTLAFK 181
 118 GKYYVYLDMAHADNV 133
 182 PAKELFSDVEFHYENTL 197

RESULT 8
 048545 PRELIMINARY; PRT; 1946 AA.

01-NOV-1996 (TREMBLrel. 01, Created)
 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 01-OC-2000 (TREMBLrel. 15, Last annotation update)
 PROTEINASE (PRT) PRECURSOR (PRTB).
 S Lactococcus delbrueckii.
 C Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
 C Lactobacillus
 X NCBI_Taxid:1584;

SEQUENCE FROM N.A.
 MEDLINE-96236017; PubMed-8655480;
 A Gilbert C., Atlan D., Blanc B., Portaller R., Germond J.E.,
 Lapiere L., Mollet B.;
 "A new cell surface proteinase: sequencing and analysis of the prtb
 gene from Lactobacillus delbrueckii subsp. bulgaricus";
 J. Bacteriol. 178:3059-3065(1996).
 R EMBL: L48487; AAC41529.1;
 R HSP: P00782; ISUE.
 R Interpro: IPR00209;
 R Interpro: IPR003137;
 R Pfam: PF00082; Peptidase_S8; 3.
 R Pfam: PF02225; PA; 1.
 R PRINTS: PR00723; SUBTILISIN.
 R PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
 R PROSITE: PS00138; SUBTILASE_SER; 1.
 SIGNAL
 SIGNAL 1 34 POTENTIAL.
 CHAIN 193 1946 PROTEINASE.
 SEQUENCE 1946 AA; 212315 MW; 21EF17D02E79C6A0 CRC64;

Query Match 11.5%; Score 82; DB 2; Length 1946;
 Best Local Similarity 23.5%; Pred. No. 11e+02;
 Matches 32; Conservative 17; Mismatches 31; Indels 56; Gaps 5;

28 VDSQSSQKSNLTP-----DQVSQKEGIAQEQ-----IVIKITDQGYTSHG 70
 171 IDVKDLPQVKNTVPKYVHPNDSADQMAQVQWQDQKKEGSAVISITIDIGDSSHQ 230
 71 D-----HYHYNGKVPDALFSEELMKDPYOLKADIYV 106
 231 DKLKDGVSATLSKSEVSDSKSLGKGKYTEKVPYGY-----NYDKNGQIYD 279
 107 -----EVKGGYIIVDQ 118

DB 280 NGCGEMHGOHVAGING 295

RESULT 9

ID 016944 PRELIMINARY; PRT; 703 AA.

AC 016944;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE PROHORMONE CONVERTASE 1.
 OS Aplysia californica (California sea hare).
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidae;
 OC Aplysiidae; Aplysia.
 X NCBI_Taxid:6500;
 RN [1]
 SEQUENCE FROM N.A.
 MEDLINE-96213914; PubMed-8639270;
 RA Gorham E.L., Nagle G.T., Smith J.S., Shen H., Kurosky A.;
 RT "Molecular cloning of prohormone convertase 1 from the atrial gland of
 Aplysia";
 RL DNA Cell Biol. 15:339-345(1996).
 DR EMBL: U04081; AAB06591.1; .
 DR HSP: Q99405; IMPT.
 DR Interpro: IPR00209;
 DR Interpro: IPR00504;
 DR Interpro: IPR002884;
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR Pfam: PF01483; P; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PRODOM: PD000717; . 1.
 DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
 DR PROSITE: PS00136; SUBTILASE_ASP; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN_1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 SEQUENCE 703 AA; 77575 MW; 208C9D7006088A38 CRC64;

Query Match 11.5%; Score 81.5; DB 5; Length 703;
 Best Local Similarity 20.9%; Pred. No. 35;
 Matches 31; Conservative 21; Mismatches 39; Indels 57; Gaps 4;

14 HROQENKNNRVYVDSQSSQKSE----- 38
 DB 86 HHTRKLSSEDRVAFVQDQQRKRVKGLVDELDRELHRELAREIAGGELHDELHEWT 145
 QY 39 -NLPQVSG-----KEGIAQEQIVIKITDQGYTSHGDHYHYNGKVPYD 83
 DB 146 LNPTEGEVSRSDEVRADLGKAVMKKGTGKIVITLDDGIERTHPDLKSNYDPEASVD 205

QY 84 ALFSEELMKDPYOLKADIYNEVKG 111
 DB 206 --FNDEDEDPSPRY-----DITENKRG 226

RESULT 10

ID 016971 PRELIMINARY; PRT; 712 AA.

AC 016971;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE PCIB.
 GN PCIB.
 OS Aplysia californica (California sea hare).
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidae;
 OC Aplysiidae; Aplysia.
 X NCBI_Taxid:6500;
 RN [1]
 SEQUENCE FROM N.A.
 MEDLINE-94213751; PubMed-8161455;
 RX Chun J.Y., Korner J., Kreiner T., Scheller R.H., Axel R.;

RT The function and differential sorting of a family of aplysia
 RT prothormone processing enzymes.
 RL Neuron 12:831-844(1994).
 DR EMBL: L28767; AAA27768.1;
 DR HSSP: 099405; IMPT.
 DR InterPro: IPR000209;
 DR InterPro: IPR000504;
 DR InterPro: IPR002884;
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR Pfam: PF01483; P; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PRODOM: PD000717; -; 1.
 DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
 DR PROSITE: PS00136; SUBTILASE_ASF; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN_1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 SO SEQUENCE 712 AA; 78526 MW; C64957F3C61E860F CRC64;

Query Match 11.4%; Score 81; DB 5; Length 712;
 Best Local Similarity 20.8%; Pred. No. 39;
 Matches 31; Conservative 21; Mismatches 39; Indels 58; Gaps 4;

QY 14 HRSQENKDNKRVSYVDGSSQSKSE----- 38
 DB 94 HNRKRLSEERVAFEVQOQKRKRVKGLVEDRELHRELAETAAAGGELHDELHHEM 153
 QY 39 --NLTPDOVS-----REGIOAEQIVIKITDQGVTSGHGHYHYNGKRVY 82
 DB 154 YLNPITSEVSRSEVADLVAVRWKKGITGKIVTITDGGIERHHPDLKSNYDEAST 213
 QY 83 DALFSEELMKDPNYQLKADIVNEVKG 111
 DB 214 D--FNDEDEPSRY-----DITENKMG 235

RESULT 11
 ID 092M97 PRELIMINARY; PRT; 567 AA.
 AC 092M97;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 GN FLAGELLAR M-RING PROTEIN.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter
 OX NCBI_TaxID=85963;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Molt D.T., King B.L., Brown E.D., Doly P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummiano P.J., Caruso A., Ulla-Nickelsen K., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voyls G.F.,
 RA Trust T.J.;
 RT Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori.
 RL Nature 397:176-180(1999).
 DR EMBL: AE001468; AAC05899.1;
 DR InterPro: IPR000067;
 DR InterPro: IPR002920;
 DR Pfam: PF01514; YscJ; 1.
 DR PRINTS: PR01009; FLGMRINGFLIF.
 SO SEQUENCE 567 AA; 63563 MW; 458B9BEF2C85C824 CRC64;

Query Match 11.2%; Score 79.5; DB 2; Length 567;
 Best Local Similarity 23.9%; Pred. No. 40;
 Matches 38; Conservative 22; Mismatches 52; Indels 47; Gaps 7;

QY 10 ALNHSQENKDN-----NRV-----SYVDSQSSQSKSENLTPDY-- 45

DB 235 ALEQLRYKQNFENLEKNTLVNLIAPVGGKKNVAVRVNAEFDSQKSTKEFFDNNVVR 294
 QY 46 -----SQREGIOAEQIVIKITDQGVTSGHGHYHYNGKRVYDALFSEELMKDPNYQL 99
 DB 295 SQONLEEKKEGAPKQVQV--GVPGVSNIGPVQGLKDKKEE---KYKSQNTTNEV 347
 QY 100 KADIVNEVKGXY-----IIRYDGKRYVYLKDAHA 130
 DB 348 --GKTISELKGEGCLVRLNAVVVDGKIKKIALDEGANA 384

RESULT 12
 ID 096185 PRELIMINARY; PRT; 2573 AA.
 AC 096185;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE HYPOTHETICAL 308.1 KDA PROTEIN.
 GN PF80460C.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99021743; PubMed=9804551;
 RA Gardner M.J., Tetteilin H., Carucci D.J., Cummings L.M., Aravind L.,
 RA Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,
 RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
 RA Salberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
 RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
 RT Chromosome 2 sequence of the human malaria parasite Plasmodium
 RT falciparum.
 RL Science 282:1126-1132(1998).
 DR EMBL: AE001396; AAC71881.1;
 KW Hypothetical protein.
 SO SEQUENCE 2573 AA; 308142 MW; 49E0D7E8967066CF CRC64;

Query Match 11.2%; Score 79.5; DB 5; Length 2573;
 Best Local Similarity 20.8%; Pred. No. 2,5e+02;
 Matches 31; Conservative 25; Mismatches 38; Indels 55; Gaps 6;

QY 12 NQRSQENKDNKRVSYVDGSSQSKSE-----NLTPD-----QY 45
 DB 2421 NQNN 2480
 QY 46 SQREGIOAEQIVIKITDQGVTSGHG---DHYH-----YNGKVPYDA 84
 DB 2481 TCKEVEVNEIIOKTKRRF--HNIELEKHYCYDLFKRKLENTYRNTYKRRKILIN 2537
 QY 85 LFESEELMKDPNYQLKADIVNEVKGXY 113
 DB 2538 LITNKNI-----FYKEHDIYVNVKIOFI 2561

RESULT 13
 ID 095937 PRELIMINARY; PRT; 264 AA.
 AC 095937;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE LIPASE-ESTERASE (EC 3.1.1.3) (TRIGLYCEROL LIPASE) (TRIGLYCERIDE
 DE LIPASE) (TRIBUTYRASE).
 GN LIPI.
 OS Mycoplasma mycoides.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Entomoplasmataceae.
 OX NCBI_TaxID=2102;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-Y-GOAT:
 RX MEDLINE-95309706; PubMed-7789792;
 RA Rawadi G., Lalanne J.L., Roulland-Dussoix D.;
 RT Cloning and characterization of the lipase operon from Mycoplasma
 RL mycoides subspecies mycoides LC.";
 CC Gene 158:107-111(1995)
 CC -1- CATALYTIC ACTIVITY: TRIACYLGLYCEROL + H(2)O -> DIACYLGLYCEROL + A
 CC PARTIAL ACID ANION.
 CC -1- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
 DR EMBL: 017036; AAA5964.1;
 DR InterPro: IPR000073;
 DR InterPro: IPR000379;
 DR Pfam: PF00561; abhydrolase_1.
 KW Hydrolase.
 SQ SEQUENCE 264 AA; 31385 MW; 3367814A9A807307 CRC64;

Query Match 11.0%; Score 78; DB 2; Length 264;
 Best Local Similarity 26.4%; Pred. No. 22;
 Matches 29; Conservative 15; Mismatches 50; Indels 16; Gaps 4;

2Y 9 VALNGHRSQENKNNKVSVDGSSQSKSENTLPDVSQKEGIAEQIYV--KITDQGY 66
 2B 5 YQYNYVKKNNNDNENIIVHYGNS-----PREFELKNIQDDQIMHNFQDIYV 57
 67 TSHGHYHYNKRVFYDALFSEELMKDNPYQLKADYNEVKGYYIIV 116
 58 KFYVKH-----KATVEG-FAGLLIHFEONQIKNVVAIGHSGGVISI 100

RESULT 14
 ID 046298 PRELIMINARY; PRT; 438 AA.
 NC 046298;
 NC 01-NOV-1996 (TREMBLrel. 01, Created)
 NC 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 NC 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 NC SPORE CORTEX-LYTIC ENZYME PRE-PRO-FORM.
 NC SLOC.
 NC Clostridium perfringens.
 NC Bacteria; Firmicutes; Bacillae/Clostridium group; Clostridaceae;
 NC Clostridium.
 NC NCBI_Taxid-1502;
 NC [1]
 NC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 NC STRAIN-S40;
 NC MEDLINE-96036223; PubMed-7582025;
 NC Miyata S., Moriyama R., Miyahara N., Makino S.;
 NC "A gene (slec) encoding a spore-cortex-lytic enzyme from Clostridium
 NC perfringens S40 spore; cloning, sequence analysis and molecular
 NC characterization.";
 NC Microbiology 141:2643-2650(1995).
 NC -1- FUNCTION: THE ENZYME IN THE DORMANT SPORE HAS NO CORTEX-LYTIC
 NC ACTIVITY
 NC -1- SUBCELLULAR LOCATION: THE PRO-ENZYME IS NON-COVALENTLY ATTACHED TO
 NC THE EXTERIOR OF THE CORTEX LAYER, AND THE PROFORM MAY BE PROCESSED
 NC TO RELEASE THE ACTIVE ENZYME.
 NC EMBL: D45024; BAA08081.1;
 NC HSP: P00733; ILBU.
 NC InterPro: IPR001899;
 NC InterPro: IPR002477;
 NC Pfam: PF01471; PG_binding_1;
 NC Pfam: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 NC PROPEP 1 114
 NC PROPEP 1 114
 NC CHAIN 115 438 SPORE CORTEX-LYTIC ENZYME PRO-FORM.
 NC PROPEP 115 149
 NC CHAIN 1 149
 NC CHAIN 150 438 SPORE CORTEX-LYTIC ENZYME.
 NC DOMAIN 43 112 MOTIF B.
 NC DOMAIN 328 400 MOTIF A.
 NC SEQUENCE 438 AA; 49639 MW; F9A4D0D1EBEF101 CRC64;

Query Match 11.0%; Score 78; DB 2; Length 438;
 Best Local Similarity 28.6%; Pred. No. 40;
 Matches 26; Conservative 10; Mismatches 27; Indels 28; Gaps 4;

OY 58 KITDQGYTSHGDHYHYNGVPPDAL--FSEELAMD-----PNYQLKDA- 102
 DB 278 INVQCPGWMTQMSKRYLGDGDEGYPYDILTSFYDDELKSAKRVGSPRSYGYTLKGY 337
 OY 103 -----DIYNEVKGX-----IKYDQGY 120
 DB 338 SGEPYRVIOEQLMAISRAYPLIPKIAVDGKT 368

RESULT 15
 ID 09PHW6 PRELIMINARY; PRT; 642 AA.
 NC 09PHW6
 AC 09PHW6;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE PUTATIVE FLAGELLAR HOOK-ASSOCIATED PROTEIN.
 GN FLUJ OR C30548.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 OC Campylobacter.
 OC NCBI_Taxid-197;
 NC [1]
 NC SEQUENCE FROM N.A.
 NC STRAIN-NCRC 11168;
 NC MEDLINE-20150912; PubMed-10686204;
 NC Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
 NC Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,
 NC Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 NC Quail M.A., Rance M.A., Rutherford K.M., Van Vleet A.H.M.,
 NC Whitehead S., Barrett B.G.;
 NC "The genome sequence of the food-borne pathogen Campylobacter jejuni
 NC reveals hypervariable sequences";
 NC Nature 403:665-668(2000).
 NC DR EMBL: AL139075; CAB75184.1;
 NC DR InterPro: IPR003481;
 NC Pfam: PF02465; FliD; 1
 NC SEQUENCE 642 AA; 69775 MW; 40F20DA1668EFA3 CRC64;

Query Match 11.0%; Score 78; DB 2; Length 642;
 Best Local Similarity 28.3%; Pred. No. 63;
 Matches 28; Conservative 16; Mismatches 39; Indels 16; Gaps 4;

OY 21 DNNKVSVDGSSQSKSENTLPDVSQKEGIAEQIYVKITDQGYVTSH---GDHYHY 76
 DB 84 DNPASLTVNGSVALQSNINIVTQALQADYVQSGKLA---NDGGFVNAQNLGTADLTFPS 140
 OY 77 NGKVPYDALFSEELMKDNPYQLKADYNEVKGYYI 114
 DB 141 NGK-----EYTVVDKNTTYRDLADKINEASGEIV 171

Search completed: September 26, 2001, 22:33:48
 Job time: 267 sec

ALL INFORMATION CONTAINED
HEREIN IS UNCLASSIFIED
DATE 09-27-01 BY 60324
UCBAW

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 22:31:17 ; Search time 32.54 seconds
(without alignments)
86.057 Million cell updates/sec

Title: us-09-471-255-2_copy_15_150

Perfect score: 711

Sequence: 1 IVSLICAYALNQRSQENK.....DGKYYVYLKDAHADNVRTK 136

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents AA:*

- 1: /cgn2.6/ptodata/2/laa/5A_COMB.pep.*
- 2: /cgn2.6/ptodata/2/laa/5B_COMB.pep.*
- 3: /cgn2.6/ptodata/2/laa/6A_COMB.pep.*
- 4: /cgn2.6/ptodata/2/laa/6B_COMB.pep.*
- 5: /cgn2.6/ptodata/2/laa/PCTUS_COMB.pep.*
- 6: /cgn2.6/ptodata/2/laa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| No. | Score | Match | Length | ID | Description |
|-----|-------|-------|--------|--------------------|-------------------|
| 1 | 663 | 93.2 | 447 | US-08-961-083-182 | Sequence 182, App |
| 2 | 575.5 | 80.9 | 763 | US-08-961-083-66 | Sequence 66, Appl |
| 3 | 574 | 80.7 | 796 | US-08-961-083-56 | Sequence 56, Appl |
| 4 | 71.5 | 10.1 | 289 | US-08-284-941-10 | Sequence 10, Appl |
| 5 | 71.5 | 10.1 | 289 | US-08-447-642-10 | Sequence 10, Appl |
| 6 | 71.5 | 10.1 | 289 | PCT-US93-02147A-10 | Sequence 10, Appl |
| 7 | 71.5 | 10.1 | 753 | US-08-712-241-3 | Sequence 3, Appl1 |
| 8 | 71.5 | 10.1 | 753 | PCT-US92-10621-3 | Sequence 3, Appl1 |
| 9 | 71.5 | 10.1 | 753 | PCT-US94-02233-3 | Sequence 3, Appl1 |
| 10 | 69.5 | 9.8 | 208 | US-08-896-933-30 | Sequence 30, Appl |
| 11 | 69.5 | 9.8 | 208 | US-08-480-604A-21 | Sequence 21, Appl |
| 12 | 69.5 | 9.8 | 608 | US-08-405-496A-21 | Sequence 21, Appl |
| 13 | 69.5 | 9.8 | 609 | US-08-480-604A-10 | Sequence 10, Appl |
| 14 | 69.5 | 9.8 | 2366 | US-08-480-604A-30 | Sequence 30, Appl |
| 15 | 69.5 | 9.8 | 2366 | US-08-405-496A-10 | Sequence 10, Appl |
| 16 | 68.5 | 9.6 | 753 | US-08-712-241-2 | Sequence 2, Appl1 |
| 17 | 68 | 9.6 | 178 | US-08-650-528-2 | Sequence 2, Appl1 |
| 18 | 68 | 9.6 | 178 | US-09-060-584-2 | Sequence 2, Appl1 |
| 19 | 68 | 9.6 | 178 | US-09-413-140A-2 | Sequence 2, Appl1 |
| 20 | 67.5 | 9.5 | 519 | US-08-997-445D-2 | Sequence 2, Appl1 |
| 21 | 66.5 | 9.4 | 203 | US-08-801-740-8 | Sequence 8, Appl1 |
| 22 | 66.5 | 9.4 | 203 | US-08-801-740-8 | Sequence 8, Appl1 |
| 23 | 66 | 9.3 | 248 | US-09-067-800-6 | Sequence 6, Appl1 |
| 24 | 65.5 | 9.2 | 520 | US-08-961-083-122 | Sequence 122, App |
| 25 | 65.5 | 9.2 | 775 | US-07-603-133B-14 | Sequence 14, Appl |
| 26 | 65.5 | 9.2 | 923 | US-08-936-135-6 | Sequence 6, Appl1 |
| 27 | 65.5 | 9.2 | 1040 | US-08-961-083-118 | Sequence 118, App |

Issued Patents

| | | | | | | |
|----|------|-----|------|---|-------------------|--------------------|
| 28 | 65.5 | 9.2 | 1528 | 1 | US-08-326-117A-2 | Sequence 2, Appl1 |
| 29 | 65.5 | 9.2 | 1528 | 3 | US-08-982-129-2 | Sequence 2, Appl1 |
| 30 | 65 | 9.1 | 378 | 4 | US-08-975-762-70 | Sequence 70, Appl |
| 31 | 64.5 | 9.1 | 382 | 2 | US-08-504-265B-75 | Sequence 75, Appl |
| 32 | 64.5 | 9.1 | 798 | 1 | US-08-190-802A-64 | Sequence 64, Appl |
| 33 | 64.5 | 9.1 | 798 | 1 | US-08-190-802A-68 | Sequence 68, Appl |
| 34 | 64.5 | 9.1 | 798 | 2 | US-08-308-818-2 | Sequence 2, Appl1 |
| 35 | 64.5 | 9.1 | 798 | 2 | US-07-728-215-30 | Sequence 30, Appl |
| 36 | 64 | 9.0 | 462 | 1 | US-08-267-092A-1 | Sequence 1, Appl1 |
| 37 | 64 | 9.0 | 462 | 2 | US-08-540-412-1 | Sequence 1, Appl1 |
| 38 | 64 | 9.0 | 462 | 4 | US-09-051-342-1 | Sequence 1, Appl1 |
| 39 | 64 | 9.0 | 462 | 4 | US-08-468-161-1 | Sequence 1, Appl1 |
| 40 | 64 | 9.0 | 462 | 4 | US-09-051-759-1 | Sequence 1, Appl1 |
| 41 | 64 | 9.0 | 462 | 5 | PCT-US95-08156-1 | Sequence 1, Appl1 |
| 42 | 64 | 9.0 | 522 | 6 | RE34608-6 | Patent No. RE34,60 |
| 43 | 64 | 9.0 | 560 | 2 | US-09-132-619-10 | Sequence 10, Appl |
| 44 | 64 | 9.0 | 560 | 3 | US-09-282-803B-10 | Sequence 10, Appl |
| 45 | 64 | 9.0 | 1112 | 2 | US-08-714-402-2 | Sequence 2, Appl1 |

ALIGNMENTS

RESULT 1
US-08-961-083-182
; Sequence 182, Application US/08961083
; Patent No. 6159463
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 182:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-182

Query Match 93.2%; Score 663; DB 4; Length 447;

Best Local Similarity 100.0%; Pred. No. 1.8e-66;

Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LNHQRSEKNNRVSYDGSOSKSENLPDVSQKESIQAEQIVKITDQGYVTSHG 70

Db 1 LNHRSQENKNNRVSVDGSSQSKSENLPDQVSKQEGIAEQIVIKITDQGYTSHG 60
QY 71 DHHYHNGKVPYDALFSEELMKDPNYQLKADIVNEVKGYYIIKVDKYYVYLKDAHA 130
Db 61 DHHYHNGKVPYDALFSEELMKDPNYQLKADIVNEVKGYYIIKVDKYYVYLKDAHA 120
QY 131 DNVRTK 136
Db 121 DNVRTK 126

RESULT 2

US-08-961-083-66
Sequence 66, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-66

Query Match 80.9%, Score 575.5; DB 4; Length 763;
Best Local Similarity 78.6%, Pred. No. 2.7e-56;
Matches 103; Conservative 18; Mismatches 9; Indels 1; Gaps 1;

QY 7 CAYALNQRS-OENKNNRVSVDGSSQSKSENLPDQVSKQEGIAEQIVIKITDQGY 65
Db 1 CSEYELGRHQAGVKNRSNYSIDGDAQKAKENLTPDEVSKREGINAEQIVIKITDQGY 60
QY 66 VHSQDHHYHNGKVPYDALFSEELMKDPNYQLKADIVNEVKGYYIIKVDKYYVYLK 125
Db 61 VHSQDHHYHNGKVPYDALFSEELMKDPNYQLKADIVNEVKGYYIIKVDKYYVYLK 120
QY 126 DAHADNVRTK 136
Db 121 DAHADNVRTK 131

RESULT 3

US-08-961-083-56
Sequence 56, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 796 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-56

Query Match 80.7%, Score 574; DB 4; Length 796;
Best Local Similarity 80.6%, Pred. No. 4.3e-56;
Matches 104; Conservative 16; Mismatches 9; Indels 0; Gaps 0;

QY 8 AYALNQRSQENKNNRVSVDGSSQSKSENLPDQVSKQEGIAEQIVIKITDQGYT 67
Db 1 SYELGLYQARFYKENRVSVDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYT 60
QY 68 SHGSHYHNGKVPYDALFSEELMKDPNYQLKADIVNEVKGYYIIKVDKYYVYLKDA 127
Db 61 SHGSHYHNGKVPYDALFSEELMKDPNYQLKADIVNEVKGYYIIKVDKYYVYLKDA 120
QY 128 AHADNVRTK 136
Db 121 AHADNVRTK 129

RESULT 4
US-08-284-941-10
Sequence 10, Application US/08284941
Patent No. 5863756

GENERAL INFORMATION:
APPLICANT: BARR, PHILIP J
APPLICANT: KIEFER, MICHAEL C
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND
TITLE OF INVENTION: PACE 4.1 GENE AND POLYPEPTIDES IN CELLS
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: COOLEY GODWARD CASTRO, HUDDLESON & TATUM
STREET: FIVE PALO ALTO SQUARE

ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/02233
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 748P3PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 753 amino acids
TYPE: amino acid
TOPOLOGY: linear
PCT-US94-02233-3

Query Match 10.1%; Score 71.5; DB 5; Length 753;
Best Local Similarity 31.2%; Pred. No. 13;
Matches 20; Conservative 11; Mismatches 26; Indels 7; Gaps 2;

QY 48 KEGIOAIOYIKITDGYVSHGDHYHYNGKVPYDALFSEELMKDPNTQLDADIYNE 107
DB 153 EKGITGKGVYTVLDDLENNHTDIYANYDPEASYD--FNDNDHDPFRY-----DLTNE 205
108 VKGG 111

DB 206 NKGK 209

RESULT 10
US-08-896-933-30
Sequence 30, Application US/08896933
Patent No. 6221351
GENERAL INFORMATION:
APPLICANT: Terman, David S.
TITLE OF INVENTION: TUMOR KILLING EFFECTS OF ENTEROTOXINS,
FILE REFERENCE: 09629/005002
CURRENT APPLICATION NUMBER: US/08/896,933
CURRENT FILING DATE: 1997-07-18
EARLIER APPLICATION NUMBER: 08/252,978
EARLIER FILING DATE: 1994-06-02
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 30
LENGTH: 208
TYPE: PRT
ORGANISM: Streptococcus pyogenes
US-08-896-933-30

Query Match 9.8%; Score 69.5; DB 4; Length 208;
Best Local Similarity 22.6%; Pred. No. 3.3;
Matches 33; Conservative 25; Mismatches 47; Indels 41; Gaps 8;

QY 9 YALNDR-----SGENKDNRY---STYDGSOSSCKSENTLPDOVSQREGIOAQ 55
DB 76 YILNSTGEIYIGTIPPAONKNVNRKILGNLTISC-----ESQONLNKLTLELDIYTFQE 131
QY 56 IYIKITDGYVSHGDHYHYNGKVPY-----DALFSEELMKDPNTQLDADIY 105
DB 132 IDEFKI--RYILM--DNRYITDTSYSGRLEIGTKDNHEIDLPDSNSTR--SDIF 185
QY 106 NEYKGGYIIRKDGKRYVYLKDAHAD 131
DB 186 AKTKDNRIIN-----MKNSHFD 203

RESULT 11
US-08-480-604A-21
Sequence 21, Application US/08480604A
Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
APPLICANT: Terman, David S.
TITLE OF INVENTION: TUMOR KILLING EFFECTS OF ENTEROTOXINS,
FILE REFERENCE: 09629/005002
CURRENT APPLICATION NUMBER: US/08/896,933
CURRENT FILING DATE: 1997-07-18
EARLIER APPLICATION NUMBER: 08/252,978
EARLIER FILING DATE: 1994-06-02
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 30
LENGTH: 208
TYPE: PRT
ORGANISM: Streptococcus pyogenes
US-08-896-933-30

Query Match 9.8%; Score 69.5; DB 4; Length 208;
Best Local Similarity 22.6%; Pred. No. 3.3;
Matches 33; Conservative 25; Mismatches 47; Indels 41; Gaps 8;

MOLECULE TYPE: protein
US-08-480-604A-21

Query Match 9.8%; Score 69.5; DB 1; Length 608;
Best Local Similarity 22.1%; Pred. No. 16;
Matches 27; Conservative 26; Mismatches 42; Indels 27; Gaps 6;

QY 16 SOENKNN-RVSYV-----DGSOSOKSENLPDQVSKGIGQAEQIVIKT-----61
DB 1 SEENKVSQKIRVNFVKDITLANKLSFN-----SDKQDVPVSEIILSFTPSYEDGLI 55
QY 62 --DGGVTSHGDIHYHNGKVPDADFSEELLKMDPNYOLKADIVNEVKGYYIIVDGK 119
DB 56 GYDLGLVSLYNEKFYINN-----FGMAYSGLIYINDSLYFKRP--VNNLTGFTVVGDDK 109
QY 120 YY 121
DB 110 YY 111

RESULT 12

US-08-405-496A-21
Sequence 21, Application US/08405496A
Patent No. 5919665

GENERAL INFORMATION:

APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
TITLE OF INVENTION: NEOROTOXIN
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 608 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown

MOLECULE TYPE: protein
US-08-405-496A-21

Query Match 9.8%; Score 69.5; DB 2; Length 608;
Best Local Similarity 22.1%; Pred. No. 16;
Matches 27; Conservative 26; Mismatches 42; Indels 27; Gaps 6;

QY 16 SOENKNN-RVSYV-----DGSOSOKSENLPDQVSKGIGQAEQIVIKT-----61
DB 1 SEENKVSQKIRVNFVKDITLANKLSFN-----SDKQDVPVSEIILSFTPSYEDGLI 55
QY 62 --DGGVTSHGDIHYHNGKVPDADFSEELLKMDPNYOLKADIVNEVKGYYIIVDGK 119
DB 56 GYDLGLVSLYNEKFYINN-----FGMAYSGLIYINDSLYFKRP--VNNLTGFTVVGDDK 109
QY 120 YY 121
DB 110 YY 111

RESULT 13

US-08-480-604A-30
Sequence 30, Application US/08480604A
Patent No. 5736139

GENERAL INFORMATION:

APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADAYE, NISHA V.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 609 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-30

Query Match 9.8%; Score 69.5; DB 1; Length 609;
Best Local Similarity 22.1%; Pred. No. 16;
Matches 27; Conservative 26; Mismatches 42; Indels 27; Gaps 6;

16 SQENKDN-RVSYV----DGSQSSQKSENTPPOVQKESIOAEQVIKTI-----61
DB 2 SEENKYSQVIRIRFVNFKTKLANKLSFNF-----SDKQDVPVSEIILSTPSTYEDGLI 56
QY 62 --DQGYTSHGDHYHYNGKVPYDALFSEELKMKDPYOLKADIVNEVGYYIIKVDGR 119
DB 57 GYDLGLVSLYNEKFIYNN----FGMAVSGLIYINDSLIYKRP--VNNLTGTGYTGDDK 110
QY 120 YY 121
DB 111 YY 112

RESULT 14
US-08-480-604A-10
Sequence 10, Application US/08480604A
Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPND-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-10

Query Match 9.8%; Score 69.5; DB 1; Length 2366;
Best Local Similarity 22.1%; Pred. No. 1,2e+02;
Matches 27; Conservative 26; Mismatches 42; Indels 27; Gaps 6;

16 SQENKDN-RVSYV----DGSQSSQKSENTPPOVQKESIOAEQVIKTI-----61
DB 1755 SEENKYSQVIRIRFVNFKTKLANKLSFNF-----SDKQDVPVSEIILSTPSTYEDGLI 1809
QY 62 --DQGYTSHGDHYHYNGKVPYDALFSEELKMKDPYOLKADIVNEVGYYIIKVDGR 119
DB 1810 GYDLGLVSLYNEKFIYNN----FGMAVSGLIYINDSLIYKRP--VNNLTGTGYTGDDK 1863
QY 120 YY 121
DB 1864 YY 1865

RESULT 15
US-08-405-496A-10
Sequence 10, Application US/08405496A
Patent No. 5919665
GENERAL INFORMATION:
APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.

REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-403-496A-10

Query Match 9.8%; Score 69.5; DB 2; Length 2366;
Best Local Similarity 22.1%; Pred. No. 1.2e+02;

Matches 27; Conservative 26; Mismatches 42; Indels 27; Gaps 6;

QY 16 SOEKDNN-RVSYV---DGSOSQKSENTLPQVSQKEGIAEQIVIKIT-----61

Db 1755 SEEKVSQVIRRVNPFKDKTLANKLSFN-----SDKQVPVSEITLSTPSIYEDGLI 1809

QY 62 --DQGYTSGDHVHYNGRVYPDALFSELLMKDPNYOLKADIVNEVNGGYIIRVDGR 119

Db 1810 GYDGLVSLYNEKFIYNN---FGMVSGLIYINDSLYFKPP--VNNLITGVTVGDGR 1863

120 YR 121

Db 1864 YR 1865

Search completed: September 26, 2001, 22:31:19
Job time: 1543 sec

Thu Sep 27 09:15:58 2001

us-09-471-255-2.ra1

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: September 26, 2001, 22:05:36 ; Search time 32.54 Seconds

(without alignments)
657,448 Million.cell updates/sec

Title: US-09-471-255-2

Perfect score: 5406

1 MFPSKRYTLAGSAYVLSL.....IEKRLPGEYIKNLSDPFA (3039)

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 197339 segs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

- 1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/backfill1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------|
| 1 | 2338 | 43.2 | 447 | 4 | US-08-961-083-182 |
| 2 | 1247 | 23.1 | 796 | 4 | US-08-961-083-56 |
| 3 | 1228.5 | 22.7 | 763 | 4 | US-08-961-083-66 |
| 4 | 223 | 4.1 | 1964 | 2 | US-08-790-912-3 |
| 5 | 223 | 4.1 | 2052 | 2 | US-08-790-912-3 |
| 6 | 205.5 | 3.8 | 1183 | 2 | US-08-447-031A-2 |
| 7 | 203 | 3.8 | 1848 | 4 | US-08-296-791-6 |
| 8 | 203 | 3.8 | 1848 | 5 | PCT-US95-10661A-6 |
| 9 | 200 | 3.7 | 571 | 4 | US-08-961-083-4 |
| 10 | 191 | 3.5 | 1507 | 6 | 5268270-2 |
| 11 | 186 | 3.4 | 1231 | 3 | US-08-904-263A-4 |
| 12 | 178.5 | 3.3 | 2465 | 2 | US-08-596-291-3 |
| 13 | 178.5 | 3.3 | 2465 | 3 | US-09-100-804-3 |
| 14 | 177 | 3.3 | 1702 | 4 | US-08-296-791-5 |
| 15 | 177 | 3.3 | 1702 | 5 | PCT-US95-10661A-5 |
| 16 | 175.5 | 3.2 | 2466 | 3 | US-09-080-855-12 |
| 17 | 175.5 | 3.2 | 2466 | 5 | PCT-US94-09943-2 |
| 18 | 174.5 | 3.2 | 2485 | 4 | US-09-290-640-46 |
| 19 | 173.5 | 3.2 | 1780 | 1 | US-08-769-309A-5 |
| 20 | 173.5 | 3.2 | 1780 | 3 | US-08-994-570-5 |
| 21 | 171 | 3.2 | 984 | 1 | US-08-242-932-2 |
| 22 | 171 | 3.2 | 984 | 1 | US-08-714-481-2 |
| 23 | 171 | 3.2 | 984 | 1 | PCT-US95-06111-2 |
| 24 | 170.5 | 3.2 | 1043 | 3 | US-08-928-361B-30 |
| 25 | 170.5 | 3.2 | 1721 | 3 | US-08-700-651-5 |
| 26 | 170.5 | 3.2 | 1721 | 3 | US-08-928-361B-6 |
| 27 | 169.5 | 3.1 | 1167 | 2 | US-08-589-756-2 |

| | | | | | | |
|----|-------|-----|------|---|-------------------|---------------------|
| 28 | 168.5 | 3.1 | 1435 | 2 | US-08-568-459A-4 | Sequence 4, Appl1 |
| 29 | 168.5 | 3.1 | 1435 | 2 | US-08-487-826B-4 | Sequence 4, Appl1 |
| 30 | 168.5 | 3.1 | 1566 | 2 | US-08-687-956A-23 | Sequence 23, Appl1 |
| 31 | 165.5 | 3.1 | 1007 | 4 | US-08-961-083-216 | Sequence 216, Appl1 |
| 32 | 163.5 | 3.0 | 1545 | 4 | US-08-296-791-4 | Sequence 4, Appl1 |
| 33 | 163.5 | 3.0 | 1545 | 5 | PCT-US95-10661A-4 | Sequence 4, Appl1 |
| 34 | 162.5 | 3.0 | 783 | 6 | 5231168-2 | Patent No. 5231168 |
| 35 | 162.5 | 3.0 | 1164 | 2 | US-08-589-756-1 | Sequence 1, Appl1 |
| 36 | 162 | 3.0 | 2409 | 2 | 5180808-2 | Patent No. 5180808 |
| 37 | 161.5 | 3.0 | 1529 | 2 | US-08-728-470-10 | Sequence 10, Appl1 |
| 38 | 161.5 | 3.0 | 1529 | 4 | US-08-719-641-10 | Sequence 10, Appl1 |
| 39 | 160.5 | 3.0 | 1600 | 2 | US-08-617-697-10 | Sequence 10, Appl1 |
| 40 | 159.5 | 3.0 | 1085 | 1 | US-08-431-080-28 | Sequence 28, Appl1 |
| 41 | 159.5 | 3.0 | 1085 | 2 | US-08-938-534-28 | Sequence 28, Appl1 |
| 42 | 159 | 2.9 | 1042 | 3 | US-08-928-361B-11 | Sequence 11, Appl1 |
| 43 | 159 | 2.9 | 1837 | 3 | US-08-928-361B-5 | Sequence 5, Appl1 |
| 44 | 158.5 | 2.9 | 802 | 4 | US-09-156-316-1 | Sequence 1, Appl1 |
| 45 | 155.5 | 2.9 | 2353 | 4 | US-09-377-155-33 | Sequence 33, Appl1 |

ALIGNMENTS

RESULT 1
Sequence 182, Appl1 Application US/08961082
Patent No. 6159469

GENERAL INFORMATION:

APPLICANT: Chol et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS: ADDRESS: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM: MEDIUM TYPE: Diskette, 3.50 Inch, 1.4MB storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA: APPLICATION NUMBER: US/08/961,083

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA: APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION: NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION: TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 182:

SEQUENCE CHARACTERISTICS: LENGTH: 447 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-961-083-182

Query Match 43.2% Score 2338 DB 4 Length 447;
Best Local Similarity 100.0% Pred. No. 2e-156;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

25 INORSGENKNNRVSYYDSOSSOSKSENLPPDOVSOREGICARQIVKIRIDGQVTSHG 84

Db 1 LNOHRSQENKDNKNVSYVDSQSOSQSENLTPDQVSOKEGIAQOYIKITDQYVTSHG 60
 QY 85 DHHYHNGKVPYDALFSEELLMKDPNTOLKADIVNEKGGYIIKVDGKYYVYLKDAHA 144
 Db 61 DHHYHNGKVPYDALFSEELLMKDPNTOLKADIVNEKGGYIIKVDGKYYVYLKDAHA 120
 QY 145 DNVTKEINROKOEHYDNKENVNVAVASQGRYTTNDGYVNPADIIEDTGNATV 204
 Db 121 DNVTKEINROKOEHYDNKENVNVAVASQGRYTTNDGYVNPADIIEDTGNATV 180
 QY 205 HGHYHITPKDLSASLAAKALAKKNNOPSQLSTSTSDNNTQSVAKGTSKPAK 264
 Db 181 HGHYHITPKDLSASLAAKALAKKNNOPSQLSTSTSDNNTQSVAKGTSKPAK 240
 QY 265 SENOSLKELYDSPAQRYSQSDGLVDPKATISRTPNGVAIPHGDHYHITPKSLAL 324
 Db 241 SENOSLKELYDSPAQRYSQSDGLVDPKATISRTPNGVAIPHGDHYHITPKSLAL 300
 QY 325 EKIARWPIGSGTSTVSTNAKPNEVYSSLSLSSNPSTLTSKELSSASDGYIFNPDI 384
 Db 301 EKIARWPIGSGTSTVSTNAKPNEVYSSLSLSSNPSTLTSKELSSASDGYIFNPDI 360
 QY 385 VEETATATVHGHFHITPKSNQIGOPTLPNNSLATPSLPIPIPGTSHEKHEDGYGF 444
 Db 361 VEETATATVHGHFHITPKSNQIGOPTLPNNSLATPSLPIPIPGTSHEKHEDGYGF 420
 Db 445 DANRIADESGFVMSHGDHNYEPFK 471
 Db 421 DANRIADESGFVMSHGDHNYEPFK 447

RESULT 2

US-08-961-083-56

Sequence 56, Application US/08961083

Patent No. 6159469

GENERAL INFORMATION:

APPLICANT: Chol et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESS: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,083

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 56:

SEQUENCE CHARACTERISTICS:

LENGTH: 796 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-961-083-56

Query Match 23.18; Score 1247; DB 4; Length 796;
 Best Local Similarity 41.2%; Pred. No. 1,9e-79;
 Matches 298; Conservative 95; Mismatches 180; Indels 150; Gaps 19;

QY 22 AYALNORSEKDNKNVSYVDSQSOSQSENLTPDQVSOKEGIAQOYIKITDQYV 81
 Db 1 SYELGLYQARVYKNNVSYVDSQSOSQSENLTPDQVSOKEGIAQOYIKITDQYV 60
 QY 82 SHGDHYHNGKVPYDALFSEELLMKDPNTOLKADIVNEKGGYIIKVDGKYYVYLKDA 141
 Db 61 SHGDHYHNGKVPYDALFSEELLMKDPNTOLKADIVNEKGGYIIKVDGKYYVYLKDA 120
 QY 142 AHADNVTKKEINROKOEHYDNKENVNVAVASQGRYTTNDGYVNPADIIEDT 197
 Db 121 AHADNVTKKEINROKOEHYDNKENVNVAVASQGRYTTNDGYVNPADIIEDT 180
 QY 198 GNATVHGHFHITPKDLSASLAAKALAKKNNOPSQLSTSTSDNNTQSVAKGTSKPAK 240
 Db 181 GNATVHGHFHITPKDLSASLAAKALAKKNNOPSQLSTSTSDNNTQSVAKGTSKPAK 240
 QY 249 -----NTQSVAKGTSKPAKSENOSLKELYDSPAQRYSQSDGLVDPKATISRT 301
 Db 241 VSNPCTNTNTSNNSNNSQASQSDNDLSLKQIKLPLSRHYESDGLVDPKATISRT 300
 QY 302 PNGVAIPHGDHYHITPKSLALEKIAWPI----- 333
 Db 301 ARGVAIPHGDHYHITPKSLALEKIAWPI----- 333
 QY 334 -----ISGSGTSTNAK-----PNEVYSSLSLSSNPSS 363
 Db 361 POPAPNKLIDNSSLVQVYKVEGVEEKGISRYFANDLSEYVKNLESKLISOES 420
 QY 364 -----LTSKELSSASDGYIFNPK-DIVEETATVHNG--DHHYHITPK-----SNQIG 410
 Db 421 VSHVLTAKKEVNAAPDDEFTKAYNLTLEAKALFYNKGRNSDQALDKLERLNDESTN 480
 QY 411 OPTLPNNSLATPSLPIPIPGTSHEKHE-----EDGYFDANRIIAE 452
 Db 481 KEKLYDOLLAFLAPLITHERGKNSQSEYTEDERVIAQLADKYTTSQGYIFDEHDIISD 540
 QY 453 DESGFVMSHGDHNYEPFKDITDEQIAQAOKHLE-----YTSINGDLSLSH 501
 Db 541 EGDAYVTPHMGSHWIGKDSLDKEKVAQAQYTEKGLTPSPADVAPANTG--DAAAI 599
 QY 502 EODYPGNAKKNKDKKIEKTIAGIMQYGRRESIVYKKNATIIYPHGDHHPADP--- 558
 Db 600 YNRVKG-----EKRIPLVRLPYV--EHTVYKNGNLI--PHGDHYHITPK 644
 QY 559 IDEH---KPVGIG---HSHSNYELFKPEGVAKKGNKYVTGEBLNVVILKNSTFN 611
 Db 645 FDDHYKAPNGYTLLEDPATIKYVEHDERPHSDG---WGNASEHYLAKKHSDSDPN 700
 QY 612 QNF 614
 Db 701 KNF 703

RESULT 3

US-08-961-083-66

Sequence 66, Application US/08961083

Patent No. 6159469

GENERAL INFORMATION:

APPLICANT: Chol et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESS: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

| | | | | |
|-----------------------|------------------|-------------------|------------|------------|
| Query Match | 22.7% | Score 1228.5 | DB 4 | Length 763 |
| Best Local Similarity | 40.5% | Pred. No. 3.6e-78 | | |
| Matches 296 | Conservative 101 | Mismatches 185 | Indels 149 | Gaps 24 |

RESULT 4
US-08-790-912-3
; Sequence 3, Application US/08790912

| | | | | |
|-----------------------|--------------|--------------------|-----------------|--------------|
| Query Match | 4.13; | Score 223; | DB 2; | Length 1964; |
| Best Local Similarity | 22.68; | Pred. No. 6.1e-07; | | |
| Matches 191; | Conservative | 98; | Mismatches 264; | Indels 292; |
| | | | | Gaps 48 |

QY 206 GCHHPIPEKLOLSASELAAKAHLAGNMPSOLSY---SFTASDNNQTSVANGSRSKPA 262
 Db 47 GVHHIYADSELSEEE-----KKOLVYDPIFYENDDTYLY----- 84
 QY 263 NKSENLOSLKELYDSPAQRYSISDGLVYDPAKIRTPNGVAILPBGHHTHTPIPSKLS 322
 Db 85 -YKNSNOQLAEL--PNTGSKNRRQALVAG---ASLALGLIL-----FAVSKK 128

```

QY 323 ALEKIRAMPISGTS--TVSTNAKPNVYSSLSNPSLSLTSEKLSADGYIFN 380
DB 129 VKNTVYALVAVMGNGVLYVALENNHL-----LNTN-----IDYELTSGK--LPL 176
QY 381 PKDVEETAAYIYRHDHFIYIPKSNQIQPPLPNNSLATPSPSLPINFOTSHKEED 440
DB 177 PKELSGTYIGYI--KEKETSDFEVSNQ-----EKSAATPT-----KQKV 216
QY 441 GYGFDAIRIIAEDSGFVMSGHGDNHYFFKKDLTEQIKAAQKLEVKTSNGLDLSL 500
DB 217 DYNVTPNFV---DHPSTVQAIOEQTPVSSSTKP--TEVOVVEKPESTELINPRKEKOSDS 272
QY 501 HEDDYPNAKEMDLDKIKIEKTAGIKQYGVRESIVYKREKAIYIPGDDHHADPID 560
DB 273 QEQ-----LAEHNKLETKKEEKIS-----PREKGV-----NTLNPQD 305
QY 561 E-----HKFVIGSHSNTELEFPEGVAKKESGKRYTGEELNVNVLKNSFNQNF 614
DB 306 EYLSGQANKP-----ELLYRETIIE--TKIDFOEI-----QENP 338
QY 615 TLANGQKRVSEFSPPELEKLGIM-LVKLIT-----PDGKYLEVSG 656
DB 339 DLAEGTAVR---KQEGIKGKVEIYRIFSVKKEVSEIYSTSTJASPRIVEKGT 392
QY 657 K--VFGE-----GV-----GNINAFELDQPYLPQGTFTTIAKDYPEVSYDGTFTVPT 703
QY 393 KTOYIKQOPPEGVYHKKVQSGALYEPAL-QPELP-----EAVSDKGEPEVOP--TLPE 443
DB 444 AVV-----TDKGET--EVQPSPTVVSDEKPEQVAPLPRYKGN-----I 482
QY 760 GEIR--LPIPL--NQGTTRAGNKIPTFMANAVLDNOSTIYVPLEKENDKRSIL 816
DB 483 EQVPEPPEVETKROGPEKT--EEVPV-----KPEETPVNPEGTEGTSI- 527
QY 817 POFARN---KQOENSKLDEKVEEPTSEKV-----EKEKLSSTGSTS 856
DB 528 -QOENPVQPAEESTTSEKVSPTSSSENGEVSNSDSSTSVGESKNEPHNDKSENS 586
QY 857 NSTLEVPYTPVOEKVAKAEYSKMLENVLFMDGTIELYLPDSGEVYKKNMADPTGEA 916
DB 587 EKYTEEP--VNP-----NEGIVE-----GTSNOSTEK 612
QY 917 POGNGENKPSNGV---STGTVENOPTENK--ADSLPEAPNEKPYKPSN---TDNG 967
DB 613 PVQPAEETQTSKRIANENGVEVSNKPSDSKPVESNOPEKNGTATKPNESGNTSNG 672
QY 968 MLNPE 972
DB 673 QTEPE 677

```

RESULT 5

38-790-912-2

Sequence 2, Application US/08790912

Patent No. 5976542

GENERAL INFORMATION:

APPLICANT: Welsch, Jeffrey N.

APPLICANT: Plaut, Andrew G.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: PANITCH SCHWARZ JACOBS & NADDEL, P. C.

STREET: 1601 Market Street, 36th Floor

CITY: Philadelphia

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19103-2398

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/790,912
FILING DATE: 23-JAN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,939
FILING DATE: 23-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Leary, Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 7600-401
TELEPHONE: (215) 567-2020
TELEFAX: (215) 567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2052 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-790-912-2

```

Query Match: 4.1%; Score 223; DB 2; Length 2052;

Best Local Similarity: 22.6%; Pred. No. 6.5e-07; Matches 191; Conservative 98; Mismatches 264; Indels 292; Gaps 48;

```

QY 206 GGHYHYIPKSDLSNSELAAKHAHLAKGNMOPQSLSY--SSASDNTQSVANGSTSKPA 262
DB 120 GVHYKYVADELSSEE-----KKQVLYOIPYVWEDDETYLYV----- 157
QY 263 NKSENTASLAKELYDSASQRTSSDGLVDPKAITIRTEGVAIPRGDHYHFIYPSKLS 322
DB 158 -YKINSQNOAL--PNTSKNERQALVAG--ASIALGILI-----FAVSKK 201
QY 323 ALEKIRAMPISGTS--TVSTNAKPNVYSSLSNPSLSLTSEKLSADGYIFN 380
DB 202 VKNTVYALVAVMGNGVLYVALENNHL-----LNTN-----IDYELTSGK--LPL 249
QY 381 PKDVEETAAYIYRHDHFIYIPKSNQIQPPLPNNSLATPSPSLPINFOTSHKEED 440
DB 250 PKELSGTYIGYI--KEKETSDFEVSNQ-----EKSAATPT-----KQKV 289
QY 441 GYGFDAIRIIAEDSGFVMSGHGDNHYFFKKDLTEQIKAAQKLEVKTSNGLDLSL 500
DB 290 DYNVTPNFV---DHPSTVQAIOEQTPVSSSTKP--TEVOVVEKPESTELINPRKEKOSDS 345
QY 501 HEDDYPNAKEMDLDKIKIEKTAGIKQYGVRESIVYKREKAIYIPGDDHHADPID 560
DB 346 QEQ-----LAEHNKLETKKEEKIS-----PREKGV-----NTLNPQD 378
QY 561 E-----HKFVIGSHSNTELEFPEGVAKKESGKRYTGEELNVNVLKNSFNQNF 614
DB 379 EYLSGQANKP-----ELLYRETIIE--TKIDFOEI-----QENP 411
QY 615 TLANGQKRVSEFSPPELEKLGIM-LVKLIT-----PDGKYLEVSG 656
DB 412 DLAEGTAVR---KQEGIKGKVEIYRIFSVKKEVSEIYSTSTJASPRIVEKGT 465
QY 657 K--VFGE-----GV-----GNINAFELDQPYLPQGTFTTIAKDYPEVSYDGTFTVPT 703
DB 466 KTOYIKQOPPEGVYHKKVQSGALYEPAL-QPELP-----EAVSDKGEPEVOP--TLPE 516
QY 704 SLAYKMASQTIFFPFHAGDYLRVNPQ---FAVPKGTDLVRYFDEFGHNAVLENNYKV 759
DB 517 AVV-----TDKGET--EVQPSPTVVSDEKPEQVAPLPRYKGN-----I 555
QY 760 GEIR--LPIPL--NQGTTRAGNKIPTFMANAVLDNOSTIYVPLEKENDKRSIL 816
DB 556 EQVPEPPEVETKROGPEKT--EEVPV-----KPEETPVNPEGTEGTSI- 600

```

; LENGTH: 100 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-447-031A-2

931 VSVNLLANGERAKTLDVYSETNMKYEKFDLPKYGEKRIETVYVEDHAKDYTDINGTGI 990

```
Db      :| | ||||| :| :  
991 TNRKTPGSTSATVKNMDDNNNQDKRPRTIKVELYQDGAATGKTALINESNNTHTWTG 105
```

QY 901 SGEYIKKMADEFT-GEAPOGNGENKPSNGKSTGTVEENOPTENKPADSL-PEADPNK-- 956
 DB 1051 LDEAKGGOVATYDELTAKNGITTHVDNDGNLITNTKTPPKPKPIPEPKDTP 1110
 QY 957 PVKPDNS 963
 DB 1111 PTKPDHS 1117

RESULT 7
 US-08-296-791-6
 Sequence 6, Application US/08296791
 Patent No. 6245337

GENERAL INFORMATION:

APPLICANT: St. Geme III, Joseph W.
 APPLICANT: Falkow, Stanley
 TITLE OF INVENTION: Haemophilus Adherence and Penetration
 TITLE OF INVENTION: Protein
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: United States
 ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/296,791
 FILING DATE: 25-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
 NAME: Treccartin, Richard F.
 REGISTRATION NUMBER: 31,801
 TELEPHONE: (415) 781-1989
 TELECOMMUNICATION INFORMATION:
 REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ. ID NO.: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 1848 amino acids

TYPE: amino acid
 TOPOLOGY: unknown
 US-08-296-791-6

Query Match 3.88; Score 203; DB 4; Length 1848;
 Best Local Similarity 19.08; Pred. No. 1.4e-05;
 Matches 259; Conservative 152; Mismatches 447; Indels 508; Gaps 64;

2 KFSKRYIAGAAVYVSLCAALAHNR-----QENKDNKRSYDGS 45
 DB 5 KFKNLFTAL-----TAAVLTPTTEALVADVDOYLFDFNAENKGFSGATNVE 55
 QY 46 QSSORSENTL--TPDOYSOKE--GLOAEQIYIKITDOGYV-----TSHGD--HYHYNG 92
 DB 56 VRDKKNSLSGALPENGIPMIDFSVVDKRIATILVNPQYVGVHNGVSELHFGWLNG 115
 QY 93 KVPYDALPSELKMDPNYOLKADIVNEKYGITIIIVDGKTYIYLKDAHADNR--TK 150
 DB 116 N-----MNNGNNAK--SHRDVASE-----ENRYTYVERKNPFETENVTSTFK 153
 QY 151 DEINROKE--HYKDEKRVSNVA-----VARSQRYTINDGYVNPDA----- 191
 DB 154 EEDQAKRREDIYPRDKLQTEVETAPLEASTANNKSGEYNSDY--PAVRLGSGTQF 210
 QY 192 -----DIIEDTGNAYIYVPHG----- 207

DB 211 IYKKSRYQILITLTKROGNLLRMWDVGDNLLEVNAATYGIAGTPYKVNHNNGILGF 270
 QY 208 -----HYIYIRSDLSASELA-----AAKALHAGNMPSPQ 238
 DB 271 GNSKEEHSDDPGILSODPLTNVAVLGDGSPFLYDREKGNMFLGSDYDWAGNKTSMQ 330
 QY 239 -----LSYS--TASDNTQSV--ANGSTSKPANKSENGOSLLEKELDPSA 281
 DB 331 EWNITKEPFAKTYQOISAGSLISNTQYTWQATGSGTSTTGGEPLSDTLTQCKDPNH 390
 QY 282 QRY-----SESDGLVDPDAITISRTPN-----GVAIPRGD-----HYH 314
 DB 391 GKSLITLKGSGTLNHNIDGAGGLFREGDYEVGTSDSTWKGAGVADAGKYTVKWH 450
 QY 315 FIPYSKLS-----ALBEK-----IARNAVISGTS 339
 DB 451 NPKYDLAKIGKGLTYVEGKGNKGLKVGDPGYTLKOKDANKVQAFSGVGVGRST 510
 QY 340 TVSTNAK--PNEVYSSL--GSLSSNPSL-----TSKELSSASD 375
 DB 511 LVLNDKQVDPNSIYFPGRGRLDNGNSLTFPHIRNIDGARVYNNMTTSITITGE 570
 QY 376 GYIFNPDIYEETATAYIVRGHFR-----YIPKSNQIQOPLPNNSLATPS 423
 DB 571 SLITNP-----NTTSYNIEAQDDHPLRISPIYRQLYFNQDR--SYTLKKG--ASTR 622
 QY 424 PSLPINFQTSHEKHEEDGYGDA-----NRILEDSSGFMASGDNHFFKDLDEQ 477
 DB 623 SELPONGESNEMWLYGKRTSDAKRVNWHINERNNGF-----NGVF--GESE 670
 QY 478 IKAOKHLEEVKTSNOLDLSISHEODY--PGNAKEKDLKRIEELIAGIMKOYKRE 535
 DB 671 TKAFON--GLINTYFNG--KSDQNFLLGNGNLNDL----- 704
 QY 536 SIYVKKERNAIYPHGDHHDPIDEKPYGIG-----HSHSYELFKDEGVAKT-- 586
 DB 705 -----NVEKGLFLSGRPTPHARDI-----AGISSTKDPHETENNEVVEDWYINRFKA 755
 QY 587 -----EEN-KYITGEELTNVYNLLKSTFNNOANTLNGOKRVFSFPELEKLGIML 640
 DB 756 TTMNVNTNASLISGR--NNAITISNTASN-----NAOVHIGY-----RTSGTVC 798
 QY 641 VR-----LTPDGKVLKESGKYGEGV--GNI-----ANPELDQPLPQOTFEYITASK 688
 DB 799 VRSDYTGTYCHNSNLEKALNSFPNLTNGVNLTEASFTLGAKLFG-----TIQSI 853
 QY 689 DYPEVSYDGTFTVPTSLAYKMASQITIFYPHAGDTYLRVNPQAFAPRGDALVRFDEFH 748
 DB 854 GTSQVN-----LKENSHWHLTGSNVNOLNLTNGHILNANQDANKVTTYNTLTVSLS 907
 QY 749 GNA-----YLENTKVEIKLPT-----FLNGGT-----TRTAGKRLPV 783
 DB 908 GNGSFYTVWDTNNKSNKRVVYVNSATGNFTLOVADKTGEPHHMLTLFLDASNAKRNMLEV 967
 QY 784 TFMANAYLDN-----OSTYIVEPILKEKNO-----DKRSIIIPQF 819
 DB 968 T-LANGSYDRGAMKTKLRNNGRIDLYNPEYERKNQYVDTNTTTPMDIQADAPS----- 1021
 QY 820 KRNKAQENSKLDEKVEEP-----KTSERKEKER 847
 DB 1022 -----AQSNEELIARVETVPVPPAPATSAIASQEPETRAPETAQAPAMEETNTASTTAP 1077
 QY 848 LSETGNSTSLLEFPYTPVPOEKVAKFPAESYKMLENLVFNNDGILIELXPGEYIKK 907
 DB 1078 KSDPATQOTENPNSESVS-----ETTKVAVENPQENETVAKNKQEAETEPPOGEVAKK 1132
 QY 908 NMADF-----TGEAPOGNGENKPSNGKSTGTVEENQPT-----ENKPADSLPEANER 956
 DB 1133 DQPTVEANTQNTAETQSEGTETETQ-----TAEKSEPTESVYSENGPEKTSQSIEDK 1187
 QY 957 PV-----KPNSTJDNGLNPEGNVSDP--MLDPALKEAPAVDPVOE 996
 DB 1188 VVVEKEKAKVETEETQKAOVTSKEPPKQAPAEVPPDTJNAEE 1233

RESULT 8
PCT-US95-10661A-6

Sequence 6, Application PC/TUS9510661A

GENERAL INFORMATION:
APPLICANT: Washington University, et al.
TITLE OF INVENTION: Hemophilus Adherence and Penetration Protein

NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10661A
FILING DATE: 16-AUG-1995
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Treccarlin, Richard F.
REGISTRATION NUMBER: 31,801
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ. ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1848 amino acids
TYPE: amino acid
TOPOLOGY: unknown
US95-10661A-6

Query Match 3.88; Score 203; DB 5; Length 1848;
Best Local Similarity 19.08; Pred. No. 1.4e-05;
Matches 259; Conservative 152; Mismatches 447; Indels 508; Gaps 64;

QY 2 KFSKKYIAGSAVYSLCAVYALNQRS-----QENKNNRYSYDGS 45
DB 5 KFKLNFIAL-----TVAIYALPTTEALVBDVDYQIFRDFAEKGFSGATNVE 55
QY 46 QSSORSEML--TPDQVSOKE--GIOAEQVYIKITDOGV-----TSHGD--HYHYNG 92
DB 56 VRDKKNSIGSALNGIPMIDFSYVDKRIATILVNPQYVGVKHSVGSSELFNGING 115
QY 93 KVPYDALFSEELMKDPYQIKADIVNEVKGITIIIVDKYIYLLDAHADNVR--TK 150
DB 116 N-----ANNNGAK--SHRDVSE-----ENRYIVTEKKNPFTEVNTSPTK 153
QY 151 DEINRQKE--HYKDEKVSNA-----VARSQRYTNDQYVFNPA-----191
DB 154 EEOAQKREDYMRDLKFTTEVAFLASTANNKGEYNSDY--PAFVRLGSGTQF 210
QY 192 -----DIIEDTGNAYIYPHGG-----207
DB 211 IYKGSRYQLITKEKQGNLRMDVGGDNLVGNAYITGIGCTPKYVHNNGLIGF 270
QY 208 ---HYHYIPKSDIASASELA-----AAKHLAKRNMQPSQ 238
DB 271 GNSKEHSDPGLISDPLNTYAVTIGDSGSPFLFYDRKDKWMLFGSIDFVAGYNNKSMQ 330

QY 239 -----LSTSS--TASDNTQSV--ANGSTKRPANKSENLOSILKELYDSPA 281
DB 331 EMNITKEHPAKIYQOYAGSLGNSNQTWQANGSTSTITIGGSEPLVDLGDGDKNNH 390
QY 282 QRY-----SSDGLVEDPAKIIISRTPN-----GVAIPRGD---HY 314
DB 391 GKSITLKGSLTLNNHIDQAGGLFEFGDEYVKQSTSTWKGAGVADGKYVTWKYH 450
QY 315 FIPYSKLS-----ALEEK-----IARVPIISGSGS 339
DB 451 NPKYDRILAKIGKGLVYEGKGNKNGELKVGQGYVLKQKADANNKVAQPSQVIGSGST 510
QY 340 TVSTNAK--ENEVYSSL--GSLSNPSSL-----TSKELSSASD 375
DB 511 LVLNDKQVDPNSIYFGRGRIDLNGSLTFPDRIINIDQARVYNNMTVNTITIGE 570
QY 376 GYIFNPQDIYETATAYIVRGGDHF-----YIKSNQIGQPLPNNSLAPPS 423
DB 571 SLITNP-----NTITSYVIEAQDDHPLRISIPRQLYFNQDNK--SYTYLKG--ASTR 622
QY 424 PSLFINGTSHKEHEDGYGDA-----NRIADESGFVMSGDHNYFFKDLTREQ 477
DB 623 SELPONSSESNMNLWMTSDAARVYNNHINERNANGF-----NGYF-----GSEE 670
QY 478 IKAOKHLEFYKTSNGSLDLSHSEODY--PGNAEKMDLDKIEKJAGIKQYGVARE 535
DB 671 TKATON--GLNATFNG--KSDONRELLTGCTYLNDL-----704
QY 536 SIYVNEKKNALYIPGHHDHAPIDEKRPVIG-----HSHSYELKPEEGVAK--586
DB 705 ---NVEKGLFLSGRPFPARDI-----AGISYTKPKPHTEENNEVVEDWIRNPKA 755
QY 587 ---EGN-KYITGEELTNVNLKNSFNNOHTLNGQKRVSRFPPELEKIGIML 640
DB 756 TTMNVTGNASLYSGR--NVANITSNTASN-----NAQVHIG--KTGDTVC 798
QY 641 VK-----LTPDGKYLEKYSKATFGEV--GNI-----ANFELDQYLPQGTFTKTIAS 688
DB 799 VRSDTYGTVCHNSLNSKALNSFNPTLRGNVILTEASSTLIGKANLFG-----TIOSI 853
QY 689 DYPEYSYDGTFTVPSLAYKMAQSOTIEYFPFAGDTYLRVNFQFVAPKGTALVRYFDEFH 748
DB 854 GTSQVN-----LKNSHMHLTGNSNVQNLMLTNGHILNQNQANKVTTNTLTVNSLS 907
QY 749 GNA-----YLENNYKVGEEKLPT-----PLNGT-----TRTAGNKLPY 783
DB 908 GNGSFYTWVDETNKSNKRVYVVKSGATGFTLQVADKGTGEPVHNELTLEDASNARRNILEV 967
QY 784 TEMNAVILDN-----OSTYIYEVPILEKENOT-----DKPSILPOF 819
DB 968 T-LANGSYDRCAMKTKLANNNGRIDLYNPEVEKRNQYVDTTNTTPNDIQDAB8-----1021
QY 820 KRNTAQENSKIDEVEEP-----FTSEKEREK 847
DB 1022 ---AOSNNEIARVEFPVPAPATASALASEQEPETPAPTAPAMEETWTANSTETAP 1077
QY 848 LSETGNSNSTLEVPYVDPVQERVAKFAESYGMKLENTLPNDGTIELTLPGEVYIK 907
DB 1078 KSDPATQTEENNSSEVPS-----ETTERVAENPQENETVAKNQEATEPTPQGEVAK 1132
QY 908 NMAOF-----TGEAPQNGENKPSENGKVSCTGVENOT-----ENRPADSLPEAPK 956
DB 1133 DQPTVEANTQNEATQSGKTEETO-----TAETKSEPTSEVYSSENOPEKTIVSQSTEDK 1187
QY 957 PV-----KPNSTNGMLNPEGNVSDP--MDPLLEAPVAVPOVE 996
DB 1188 VYVEKEEKAKYETETQKAPQVTSKEPKQAPAPAEVPTDTNABE 1233

RESULT 9
US-08-961-083-4
Sequence 4, Application US/08961083
Patent No. 6159469

GENERAL INFORMATION:

APPLICANT: Choi et al.
 TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 NUMBER OF SEQUENCES: 452
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/961,083

FILING DATE:

CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PB340P2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
 LENGTH: 571 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-961-083-4

Query Match 3.78; Score 200; DB 4; Length 571;

Best Local Similarity 22.6%; Pred. No. 4e-06;
 Matches 166; Conservative 87; Mismatches 222; Indels 258; Gaps 42;

```

QY 381 PROVEETAAATYTRHGDHHTYIPKSNQIOGPTLPNNSLATPPSPSLINGTSHEKHEED 440
DB 17 PKEISGTYIGYI-KESKTTSESEVSNQ-----KSSVATPT-----KQCKV 56
QY 441 GYGFDARIIAEDSGFVMSGHNNHYFFKKDLTEQOIKAAQKHEVKTSHNGIDSLSS 500
DB 57 DYNTPHFV---DHPSTVQAIOEOTPVSTKRP-TEGVVEKPFSTELINRKEKSSDS 112
QY 501 HEQYTPGNAKEMKDLKIEKTIAGIKMKGVAKRESIVNKEKNAIIYYPHGDHHPID 560
DB 113 QEQ-----LAHKMLTEKKEKIS-----PREKTV-----NTLNPQD 145
QY 561 E-----HKPVGISHSNTELEFPEGVAKKESKRYTGEELINVNVLKNSFNQNF 614
DB 146 EVLSGQINKP-----ELLYRETKME-----TKIDFDEI-----QENP 178
QY 615 TLANGQKRVSSPPELEKIKIIM-LVKLIT-----PDGKYLEKVSQ 656
DB 179 DLASGTVRV-----KQEGTLGKVEIVELFVSVKKEVSREIYSTTASPRIVEKGT 232
QY 657 K--VFGE-----GV-----GNLAFELDOPYLPGOTFKTTISKDYPEVSYDGTFTVPT 703
DB 233 KTVQIKQOPETGVEHKDVQSGAIVEPAI-QPELP-----EAVVSDKGEPEVOP-----TLPE 283
QY 704 SLAKMASQITFYFPHAGDYTLRVNPO-----FAVPRGTDALVRFDFEFGHGAALYENNYK 759
DB 284 AVV-----TDGERT--EVQPSPTVVSVDKGEPEVAVLPRTKGN-----I 322
QY 760 GEIK-LPIBPL-NQGTTRAGNKIPVTFMANNVLDNQSTYIYEVVPILEKENQTDKPSIL 816

```

```

DB 323 EQVKEPEVETKTKQOGEK--EEVPV-----KPLEETPVNPGDTEGTSI- 367
QY 817 POFKRN--KAQENSKIDEVEEPTSEK-----VEEKLSERTNST 855
DB 368 -QEAENFVQAEESTNTSEYV-SPDTSSKNTGEVSNPSSSTTSVGBSNPPEHDSKRNEN 425
QY 856 SNSTLEEVPTVDVQAEVAFASYGKLEVLNFNMDGTIELPLSGEVIKKNADTGE 915
DB 426 SEKTEEVV-VNP-----NEGIVE-----GTSNDETE 451
QY 916 APOGNGENKPSNGKV---STGVENOPTENRP--ADSLPEAPEKPEVKNENS---TDN 966
DB 452 KPVQPAEETQNSKINENNTGEVSNKPSKPEVESNPENKGTATKPEKNSGNTSEN 511
QY 967 GMLNPE---GNVSDPMLDALPEAPVADVQAELEFTASYGIGLDSVLFNMDGTIELR 1023
DB 512 GQTEPEPSNGNSTEDVTSNTSNGNEBKQNE-----LDPKRVKEEPTIELR 564
QY 1024 LPSGEVIKKNLSD 1036
DB 565 -----NVSD 568

```

RESULT 10

5268270-2

APPLICANT: Meyer, Thomas F; Halter, Roman; Pollner, Johannes

TITLE OF INVENTION: PROCESS FOR PRODUCING PROTEINS USING GRAM

NEGATIVE HOST CELLS

NUMBER OF SEQUENCES: 6

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/171,872

FILING DATE: 01-JUL-1987

SEQ ID NO: 2

LENGTH: 1507

Query Match 3.58; Score 191; DB 6; Length 1507;

Best Local Similarity 19.28; Pred. No. 7.2e-05; Lengths 410; Gaps 59;

Matches 256; Conservative 160; Mismatches 509; Indels 410; Gaps 59;

```

QY 16 VSLSL-CAYALNQRS-----QENKNNRVSYDGSQSSQKSENLTLPDQ 58
DB 11 ISISFLAALTPYSEALRPDQVDYQIREFDPEKNGKFFVGANDLSVKKRQNTI-GNA 69
QY 59 VSOKEGI---QAEQIVKITDQGYTS-----HGDHHTYNGKVPYDALFSEEL 105
DB 70 LSNVPMIDFVADVKNRIATVVDPOYAVSVKHAKEVHTFYGGYNGHNDVADKENERY 129
QY 106 MKDPNQ-----LKQADI-----VNEVK-----GCVIIRVD 131
DB 130 VEONNTEPHAMGASNLGRLEDYIMARENFVEEVADIAPTDAGGGLDYTKDKNRSSEFV 189
QY 132 ---GKYVYLKDAAHAD-----NVRTKDEI---NQKQEHVADNKNVSNVAVA--- 174
DB 190 RIGAGRLVYEKGVYHQBEGNEKGDLETEGLIGGNHNNKQYSAELEKQALSDQALTYGV 249
QY 175 -----RSQ-----GRITNDGY-----VFNP--ADIEDTGNAVTVPH 205
DB 250 LGDSGSLFADKQKNQWVFLGYDYWAGYKSKSMQSMNTYKKEFDKIRIQHNDAGTVKG 309
QY 206 GGHYHTPKSDLSASELAAKAHLAGKNQPSQSVSSTASDNT-----OSVAKGS--- 257
DB 310 NGEHHW--KTTGNSHIGSTAARLA--NNEGDNANGQVATEDNGTLLANONINGAGGLF 366
QY 258 ---TSKPAKSEBNQSLKELYDSFSAQRYSSDQVLPDAKIIISRTFNGVAIIPGDH 312
DB 367 EKGDYTYKGAN---NDITWLGAGID-----VADGKRVYQVQVKN---PAGDR 406
QY 363 YHFIPYSKLSALEKTIAR-NVPISGTG-----STVSNAPKPE-----V 350

```

RESULT 11
 US-08-904-263A-4
 Sequence 4, Application US/08904263A
 Patent No. 6015889
 GENERAL INFORMATION:
 APPLICANT: LINDAL, GUNNAR
 APPLICANT: STALHAMMAR-CARLEWALM,
 APPLICANT: MARGARETHA
 APPLICANT: STENBERG, LARS
 TITLE OF INVENTION: PROTEIN RIB, A CELL SURFACE PROTEIN THAT

```

TITLE OF INVENTION: CONFERS IMMUNITY TO MANY STRAINS OF THE GROUP B
TITLE OF INVENTION: STREPTOCOCCUS: PROCESS FOR PURIFICATION OF THE PROTEIN
TITLE OF INVENTION: REAGENT KIT AND PHARMACEUTICAL COMPOSITION
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/904,263A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: SVENSSON, LEONARD R
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 552-119P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1231 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-904-263A-4

Query Match: 3.4%; Score 186; DB 3; Length 1231;
Best Local Similarity 20.3%; Pred. No. 0.00012;
Matches 240; Conservative 145; Mismatches 416; Indels 384; Gaps 66;

OY 39 VSYDGSQSSQSKSENLPDQVSOREGIOABQIYIKITDGGYVWSHG---DHYHYNGKV 94
DB 35 ISFLGG--FTQGGONISTDTYFAAEVYSSGSAVILN-TNMTKNQNGRAVIDLYDKNGKI 91
OY 95 -PYDA-----LPSEELMKDPNY--QLKDADIVNEVGGY-TIKVDGKY----- 135
DB 92 DPLQLITLNSPLDKAQYVIRGGGVYFQPSLFTVGAASINTYVLTKTDGSPHTKPDGQVD 151
OY 136 -----VYLKDAH--ADVRRK-----DEINRK-----QEHKDKENKNSVA 172
DB 152 IINVSLEYNSALRDKIDEXKKAEDPKWDEGSRKVLISLDITDITDIDNNPKQSDIA 211
OY 173 VASGGRATTVDGVFNPADIIEDTG-NATVPRGGHYHPIPKSDLSASGLAAAKAHLAG 231
DB 212 -----NKITEVTINERLIVR-----IPAD-----KNDPRAG 238
OY 232 KNPQPSQLSYSTASDNNQSVAKGSTSKPANKSENIGSLK--ELYDPSAQRYSSESD 288
DB 239 KD-----QQVNVGETPKAEDSIGNLPDLPKGTVAFFP-VDRATPED 280
OY 289 GLVDPDKAIIKRTNG-----VAIRGHGHYHPIPKSLALEKIAIAWVPISGG 338
DB 281 -----KPAKVAVVYTDGSKDYVDVTVKRVVDRTDADKNDP-----AGKD 319
OY 339 STVSNKPKNPNVVSLSLSSNPSLSLTTSKE--LSSASDG-----YINPKDIYEE 387
DB 320 QQVNVGETP-KAEDSIGNLPDLKRGITVAETVYDTRATPDGRKAKAVVYYPGSGDIDYDV 378
OY 388 TATAYIYA-HGDHFHYIPKSNQI-----GQPTLPNNSIATPSPSLPIINPGTS 433
DB 379 TVKVVYDRTDADKNDPKAGKQDVVNGVETPKAEDSIGNLPDLPLPKGT--TVAFFTPVDATAP 436
OY 434 HEKHEDEQYGFADNRITIAEDSGFVMSHGDNHYIFPKKDLTEQOIAAOKHLEVTKISN 493

```

```

Db 437 GDK-----PAKVYVTPDGS-----KDVVDVYKVVDPRTDADKNDP 474
QY 494 GJSL-----SHEDYPRNAKEMDLKRIE-----EKIAGIMQY--GK 533
Db 475 GKQOVNVGEPRAEDSIGN--LPDLPKGTVAETPVDTAPGKPAVVVYTPDGS 531
QY 534 RESIVVAKENALITPHGDHHDAPIDEHNPVIGSHSHSYELFKPEGVAKKGNKYV 593
Db 532 DWDVYK-----VDPRTDADKNDPAGKQOVNVG-----ETPKAEDSIG----- 572
QY 594 GEEELVNVNLKST--FNNONFTLANGOK--RVSEFPELEKLGIMLVLI----- 644
Db 573 -----MLDLPKGTVAETPVDTAPGDPKPAVVVYTPD--GSKDTVDVYKVVDPRTD 645
QY 645 -----TPDGKLEVYSGV--FGEGVGNIANFELDQPYLGEQTKYITIASDIPVSDGT 688
Db 626 ADKNDPAGKQOVNVGEPRAEDSIGNL-----PDLPGST--VVA----- 663
QY 699 FTVPTSLAY--KMASOTIYEPHAGDTYLRVNPQFAPVPGTDALVRPD-----EHHGN 750
Db 664 FETPVDTAPGDPKPAVVVYTPDGSKDT-----VDVYKVVDPRTDADKNDP 710
QY 751 AYLENNYKVEIKLP-----IPKLNQGT-----TETAGNKIPVTEMANAYLDNOS 795
Db 711 AGKQOVNVGEPRAEDSIGNLPLDKGTVAETPVDTAPGDK--PAKVYVYTPDGSND 769
QY 796 TYIVVERPLEKENQDTPSLIPQFKRKAQENSKLDERV--DEPTSEKVEKEKLESG 852
Db 770 TVDVYKVVDPRTDADK-----NDPAGKQOVNVGEPRAEDSI--GNLPDL 815
QY 853 NGSNSGTLEVPYVDVQEVKAEASY--GMLENVLENMGTIELILPSGVITKMA 910
Db 816 KGTVAETPVDTAP--GDPKPAVVVYTPDGSK-----DTVDVYKVVDPRTDADKNDP 869
QY 911 DFGAEPQGN--GENKSENGKSTGTVENOPTENKPADSLP--EAPNEKPYV-----PE 961
Db 870 -----GKQOVNVGEPRAED--SIGNLPLDKGTVAETPVDTAPGDKPAVVVYTPD 923
QY 962 NS-----TDNGMLNPEG-----NVGSDPMLDPALEAPV-----DP 993
Db 924 GSKDTVDVYKVVDPRTDADKNDPAGKQOVNVGEPRAEDSIGNLPLDKGTVAETP 983
QY 994 VQ-----EKLEKTAHYGJSLDSVIFNMGTIELRLPSEGVAKKN 1033
Db 984 VDTATPGDPAKVYVYTPDGSKDTV--DVYKVVDPRTDADKND 1025

```

RESULT 12

US-08-596-291-3
Sequence 3, Application US/08596291
Patent No. 5821075

GENERAL INFORMATION:

APPLICANT: GONZ, LEONEL JORGE
APPLICANT: SARAS, JAN
APPLICANT: CLAESON-WELSH, LENA
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/596,291
FILING DATE: 09-AUG-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/115,573
FILING DATE: 01-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: L0461/7000
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
TELEX: 92-1742 EZEKIEL
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2465 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
US-08-596-291-3

```

Query Match 3.3%; Score 178.5; DB 2; Length 2465;

Best Local Similarity 18.7%; Pred No. 0.0011;

Matches 186; Conservative 152; Mismatches 369; Indels 261; Gaps 43;

```

QY 45 SOSQSENLTLP-----DOVSQREGIOABQIVIKITDQGYTSHGDHYHYNGKPYDAL 99
Db 1434 SPLSKREHYVTPQCTLSDDNAGQGEKVKKTQVMDYSEVTE-----ENT 1479
QY 100 ESEELMDPNTQL--KDAIVNEYKGGYIIRKVDKTYVYLKDAHADNVRKTD--EI 153
Db 1480 FEVKLFKNSSGFGFSRSDNLIPEQINASIYRV--KLIFACQPAESKIDVGVILTV 1537
QY 154 NKQOEHVNDKKNVSNVAVARSQGYTNDGYV--NPADIIDETGNAYIVPGHGH 210
Db 1538 NGASLGLSQOEVY-----SALRGTAPEVFLLCRPPOVYLPEIDTALITPLQSPAQ 1589
QY 211 YIPKSDLSASELAAKAHLAGRMOPQSLSYSTASDNTQSVANGSTSPRKNSENTQS 270
Db 1590 YLPNSKSS-----QPSCYEOSTSDENEMSKKCKPSRR----- 1629
QY 271 LKELIDSPSAORTSESDGLVDPKAIISRTPNCAIVPAGDHRIPIYSKLSALEKIR 330
Db 1630 -----DSYSDSGSGEDDLVTAIPANISNTWS-----SALHQTLSN 1665
QY 331 NY-----PISGTGSVSTNAKPNEYVSSLSNSPSLITTSKELSDADYIFNP 381
Db 1666 MYSQASHHEAPKSOEDTICTMFTYTPQKIPNKPREDSDNPSFLP-----PDMAPOGSIOP 1720
QY 382 KDIVETATAYIVRHGDHETIPKSNQGOPTLPNNSLATPSPSLINPTGSHKHEEDG 441
Db 1721 QS--SSASSSMDXY--HNNH-----ISEPTROEN--WPLKN-----DLEHNLDS 1760
QY 442 YQFDNRITIAEDS-----GFVMSGDHNNHYFFKDLLEBOJKAOK-----HLEVKTS 491
Db 1761 FELEVEELLTLTKSEKASIGFTVTKGNORIGCYVDVIODPKASQGRKPGDRILKV-- 1817
QY 492 HNGDLSLSHEDDYGNKAKEMKDLKRIEKLKAGIMKQGVNR-----ESIVYKKEK--NA 545
Db 1818 NDQVYNTKHTDVAYNLAASAKTYRLVYIGRPRITONTNVASFPATGKHLTKNKEELGFS 1876
QY 546 IYIPGHDDHHDAPIDEHNPVIGSHSHSYELFKPEGVAKKGNKYVYGEELTNVNLK 605
Db 1877 LCGGDSLYQVYISDINP-----RSVAIAIGNV-----QLDVIYHYVN 1915
QY 606 NSTFNQNTFLANGOKRSPSPPELEKLGIMLVKILITPGKYLEVYSGAVPDEGYN 665
Db 1916 G--VSTQGTLEEVNRALDMSL--PSLVTKATRNNDP--VVPSSK--RSASVAPAKSTKGMS 1969
QY 666 IANFELDQPYL--PGQTFYTIASKDYDEVSVDGTFTVPTSLAYKMAQOTIYEPHAGDTY 724

```

Db 1970 YSVGSCSPALTPNDSPS-TVAGEEINEIS-----YFGKCSFY 2007
 QY 725 -LAVNPOFAVPKGT-----DALVRFDFEFGNATLENNYKGEIKLP 766
 Db 2008 QINGSPMLTLPKESYIOEDDIYDSQAEVYISLDVYDEAONLNNEN-----2057
 QY 767 PKLNGGTRTAGNKI-PYTFMANAYLDNOSTIYEV---PILKRNQTDPSILPOKRN 822
 Db 2058 -----AAGSCGPGTLKANGKLSERTEDTDCGSLPFTFTATMNGCEECE 2108
 QY 823 KAOENSKLDEKVEPEKTS-----EKYEKELSETGNSNSTLEEVPTVDPV 869
 Db 2109 KYKSES-LIOKPOEKTDDEITWGNDELPIERTNHEDSKDSFLNDELAVLPV---2164
 QY 870 QERYAKFAESYGMKLENVLPNDGTIELYPSGEVTKRNADFTGEAPOGNGENKPSENG 929
 Db 2165 --KVLPSGKYTGANKLSVIVLNGLLDQIGPSKEL--ENLQ-----ELKPLDQC 2209
 QY 930 KYSTGYENOPTENKPADSLPEAPNEKPYKPNSTDN 966
 Db 2210 LI--GQTKENRRNRKYNILPYDATRVLGDDEGYIN 2244

RESULT 13

US-09-100-804-3
 Sequence 3, Application US/09100804
 Patent No. 6066472

GENERAL INFORMATION:

APPLICANT: GONEL, LEONEL JORGE
 APPLICANT: SARAS, JAN
 APPLICANT: CLAESSON-WELSH, LENA
 TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
 TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
 TITLE OF INVENTION: TYROSINE PHOSPHATASES
 NUMBER OF SEQUENCES: 34
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
 STREET: 600 ATLANTIC AVENUE
 CITY: BOSTON
 STATE: MASSACHUSETTS
 COUNTRY: USA
 ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/100, 804
 FILING DATE:

CLASSIFICATION:

Prior Application DATA:
 APPLICATION NUMBER: US 08/596,291
 FILING DATE: 09-AUG-1996
 APPLICATION NUMBER: US 08/115,573
 FILING DATE: 01-SEP-1993
 Prior Application DATA:
 APPLICATION NUMBER: PCT/US94/09943
 FILING DATE: 01-SEP-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: GATES, EDWARD R.
 REGISTRATION NUMBER: 31,616
 REFERENCE/DOCKET NUMBER: L0461/7003
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-720-3500
 TELEFAX: 617-720-2441

TELEX:

INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2465 amino acids
 TYPE: amino acid

TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-100-804-3

Query Match: 3.3%; Score 178.5; DB 3; Length 2465;
 Best Local Similarity: 18.7%; Pred. No. 0.0011;
 Matches 186; Conservative 152; Mismatches 398; Indels 261; Gaps 43;

QY 45 SOSQKSENITP-----DQYSQKEGLOAEQIVIKIDQGVYTHSHGHYHYNKRVPDAL 99
 Db 1434 SPYKREHVPTPOCTSLDQNAOQGEKRYKTYQVDYSFVTE-----ENT 1479
 QY 100 FSEELMKDQPNYOL---KQADIVNEKYGYYIIVDGYYYLLDAADAVNRKD---EI 153
 Db 1480 FEYKLEKNSGGLQFSRREDNLPEQIMNSIVY--KLEPAGPAESGKIDGVYLYK 1537
 QY 154 NROKREHYKDNENYNSNVAARSGRTYNDGYV--NPADIIETGNAYIYPHGGHYH 210
 Db 1538 NGASLKGISQOEVI-----SALRGTAPEVFLLCRPPGYLPEIDTALPLQSPAQ 1589
 QY 211 YIPKSLASSELAAVAHLAKRMQPSQLSYSTASDNTQYAKGSTKPAKSENLOS 270
 Db 1590 VLPNSSKDS-----OPSCVQSTSDENEMSKCKOCKSPSR-----1629
 QY 271 LKELVDSPAQRYSDDGVFDPKTIISRTPGVAIPHGDHYHFIYSKLSALEKRIAR 330
 Db 1630 -----DSTSDSSGGEEDLVTPANISNTWS-----SALHQTLSN 1665
 QY 331 MV-----PISGTGTVSTNAKPNVYSSLSNPSLSTSELSASDGYTFNP 381
 Db 1666 MVSQAQSHHAPKSOEDTICTMFTYYPQKIPKREFEDESNPSLP-----PDMAPGQSYOP 1720
 QY 382 KDVEETATAYIVRHGDHYFIKSNQIGOPTLPNNSLATPSSSLPYNCTSHKEHEOG 441
 Db 1721 QS--ESASSSMQY--HIHH-----ISEPTQEN--WPLKN-----DENHLED 1760
 QY 442 YGFDARIIMEDS-----GFVSHGDHNYFERKDLTEQIRAKO-----HLEEKTS 491
 Db 1761 FELEVELLITLISEKASLGFYTKGNRGICYVHDYIDQPAQSDGLAKGDRILIV--1817
 QY 492 HNGLDLSHEDQYPGNAKEMKDLKRIEIKIGIMQYQVCR---ESIVYKREK--NA 545
 Db 1818 -NDTDTNMTHTPVAANLRAASKTVRYIGRVRIQNTNVASFGNKLTKCKEELGFS 1876
 QY 546 IYPHGDHRAHADIPEKHPVIGISHSHNYELFPEEGVAKKGNKYTGELTNVYVLK 605
 Db 1877 LCGGHSLYQVYIISDINP-----RSVAIEGNL-----QLDVIHYVN 1915
 QY 606 NSTFNQNTFLANGOKRVSPFPELEKIGIMLVYLIPDGKYLEKVGKVFGEVGN 665
 Db 1916 G--VSTQMTLEEVNRLDMSL-PSLVKATRNDLP--VPPSR--RSVASAPSTKNGS 1969
 QY 666 IANFELDQPL-PQDFEYKVIASKDYEVSDGYFTVPTSLAYKMSQTLFFPHAGDY 724
 Db 1970 YSVGSCSPALTPNDSPS-TVAGEEINEIS-----YFGKCSFY 2007
 QY 725 -LAVNPOFAVPKGT-----DALVRFDFEFGNATLENNYKGEIKLP 766
 Db 2008 QINGSPMLTLPKESYIOEDDIYDSQAEVYISLDVYDEAONLNNEN-----2057
 QY 767 PKLNGGTRTAGNKI-PYTFMANAYLDNOSTIYEV---PILKRNQTDPSILPOKRN 822
 Db 2058 -----AAGSCGPGTLKANGKLSERTEDTDCGSLPFTFTATMNGCEECE 2108
 QY 823 KAOENSKLDEKVEPEKTS-----EKYEKELSETGNSNSTLEEVPTVDPV 869
 Db 2109 KYKSES-LIOKPOEKTDDEITWGNDELPIERTNHEDSKDSFLNDELAVLPV---2164
 QY 870 QERYAKFAESYGMKLENVLPNDGTIELYPSGEVTKRNADFTGEAPOGNGENKPSENG 929
 Db 2165 --KVLPSGKYTGANKLSVIVLNGLLDQIGPSKEL--ENLQ-----ELKPLDQC 2209

Page 12

05-08-296-791-5

1. The first step is to identify the problem or question that needs to be answered. This involves understanding the context and the specific requirements of the task.

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/10661A
 FILING DATE: 16-AUG-1995
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/296,791
 FILING DATE: 25-AUG-1994
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Treacartin, Richard F.
 REGISTRATION NUMBER: 31,801
 REFERENCE/DOCKET NUMBER: FP-59941/RFT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELE: 910 277299
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1702 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 PCT-US95-10661A-5

Query Match 3.38; Score 177; DB 5; Length 1702;
 Best Local Similarity 19.68; Pred. No. 0.00083;
 Matches 223; Conservative 141; Mismatches 431; Indels 342; Gaps 53;
 QY 84 GDHNYNGKVPY-----DALF-----SEB-----LAMD-NTQ-LKAD-----IV 119
 DB 236 GDAITGAGIPYKVNHNENLIGFGSKREHSPKGLISDPLTNVAVLGDSGSPLEFY 295
 QY 120 NEVKGGLIKYDGKYYLLKDAHADNRYTKDELINROKOEHVK--DNEKYNNAVARS 176
 DB 296 DREKGMFL--GSY-----DFAGYNKKSQWENIYKPEFAKTVLDKDTAGSLIG--- 344
 QY 177 QGRITNDGYFPNADILEDGNAYIYPHGHYHYIRKSDLSASELAARKHLAKKNOP 236
 DB 345 -----SNTQYMNW------TGKTSVLSNGSESLNVLFD--SSODTDSKKNHGSV-- 389
 QY 237 SOLYSSTASDN-----TOSVAKGSTSPANKSENLOSLKELYDPS 280
 DB 390 -TLNGSGTLTANNIDGAGGLFEGDEYKGTSDSTTWKAGVSVADGKTVTKVHNPK 448
 QY 281 AQRYS-SDGLVFPDPAIISRTPNGVAIPGHDHYHFIPYSKLSALEKRIAMPVPSGTS 339
 DB 449 SDRLAKIGKTLI-----VEGGENKGSLLKVGDTVILKQADANNKYKAFSGVIGSRS 504
 QY 340 TVSTN---AKPNEVVSLL--GSLSSNPSSLTSGKELSSASDGTTFNPKDIVEETATAYI 393
 DB 505 TVVLNDKQVDPNSIYFGFRGRLDANGNML-TFEHIRNIDG-----ARL 549
 QY 394 VRHGDHFHYIPKSNQIOQPLPNSLAT--PSPSLPDPGSHS-----KHEEDG----- 441
 DB 550 VNH-----NTSKTSTVTTGSLTLTDPPTITPINADPEDNPYAFRIKIGGGLYL 601
 QY 442 ---YGFANRIIADSDSGFYMSGHDNHYFFKKDLTEEOJKA-AOKHLEVKTSNGLD 496
 DB 602 NLENTTYAALFKGASTSELPKNGSESNENLYMGKTSDAKRVNMHINERM--NGFN 659
 QY 497 SLSSHEQ-----DYPGNAKEMKL-----DKTIE-----KIA 524
 DB 660 GYFGEERKNGNINLVTFKGSQONRFLTGTNMLNGDLKVEKGTLLFLSGRPTPHARDIA 719
 QY 525 GIMK-----QYGVKRESIYV-----NKEKNALIYPHGDHHDH---PIDHK 563
 DB 720 GISSTKQDQHAENNEVVEDMIRNFKATNINVTNATLISGRNVAANTSNITASDNA 779
 QY 564 PVGIGHSNTELEKPEEGVAKKGNKY-----YTGEELTNVYMLKNSFPNNQFTLAN 618
 DB 780 KVHIGY-----KAGDITCVASDITG--YVCTCTDKLSDKALNSFNATNVS 822

QY 619 GQKVSFSPPELEKIGINNLYLTPDG-----KYLEKSGKVFGBGVNIANFELDOP 674
 DB 823 GNVNLS-----GNANFVKGANLFTIGTGNQVRLNENHMLTGDSNNQNLMDGHI 878
 QY 675 YLPGQTFKTYIASD---YEVSTDGTFVTPSLAYMASQTLFFPHAGDTYLRNPOF 731
 DB 879 HLNQNDANNTVYTNVTLVNSLSGNSGFYLLTDLNSQGDGVVYTKSATGNTFTLQVADKT 938
 QY 732 AVPRGTALRYVDEPFIGNAVLENNYKVGKILPILPILNGOTTRTAGNKIPVFNNAVYL 791
 DB 939 GEP--TNEELTLEDA--SNA--TRNNLNVSIV-----GNTVDLGAMKTKLR 978
 QY 792 DNSTYIYEVPILEREKNOT-----DKPSI-----LP 817
 DB 979 NVNKRIVLYNPEVEKRNQYDTTITPNNIQADVPSVPSNNERIANVETPPPPAPATP 1038
 QY 818 QFRKNKQENSKLDER-----VEPPTSEK--VEKELSTGSGTS 856
 DB 1039 SETTEVAENSKQSKYVEKNEQDATETTAQNGEVAEAKPSVANTQTNVVAOSGSETE 1098
 QY 857 NSTLEVPYTPVQ--EKVAKFAESYGKLENVLFNMDGTIELYLPSEGVIKKNADPTGE 915
 DB 1099 ETOTTEIKETAKYKEKAKVEKAKVEK-----DEIOE 1134
 QY 916 APOGNGENKP-----SENGKYSTGVENOP-----TENKPADSL-- 949
 DB 1135 APOMASTSPQOAKPAPAKREYSTDKVEETOYQAOPOQOSTVYAAAEATSPRSKRAELETOP 1194
 QY 950 PEAPNEKPYRP--ENSTNGMLNPEGNVGSDPMLDPALEAPAV---DVOEKLE 999
 DB 1195 SEKNAPAPVPVPSKNOTENTTDOPTEREKTAKEETEKOEPQVQAQASPKROOSE 1251

Search completed: September 26, 2001, 22:31:17
 Job time: 1541 sec